Krzysztof Ciomek, Valentina Ferretti and Milosz Kadzinski

Predictive analytics and disused railways requalifcation: insights from a Post Factum Analysis perspective

Item type
Article (Accepted version)

Original citation:

Ciomek, Krzysztof, Ferretti, Valentina and Kadzinski, Milosz (2017). *Predictive analytics and disused railways requalifcation:insights from a Post Factum Analysis perspective.* Decision Support Systems. ISSN 0167-9236

DOI: 10.1016/j.dss.2017.10.010

Predictive analytics and disused railways requalification: insights from a Post Factum Analysis perspective

Krzysztof Ciomek[a], Valentina Ferretti[b], Miłosz Kadziński[*,a]

[a]Institute of Computing Science, Faculty of Computing, Poznań University of Technology, Piotrowo 2, 60-965 Poznań, Poland
[b]Department of Management, London School of Economics and Political Science, Houghton Street, London WC2 2AE, UK

Abstract

Strategic decision making problems in the public policy domain typically involve the comparison of competing options by different stakeholders. This paper considers a real case study oriented toward ranking potential actions for the regeneration of disused railways in Italy. The study involves multiple conflicting criteria such as an expected duration of construction works, costs, a number of potential users, and new green areas. Within this context, we demonstrate that Post Factum Analysis (PFA) coupled with Decision Aiding supports the development of robust recommendations. The role of PFA is to highlight how the actions' performances need to be modified so that the recommendation is changed in a desired way. In particular, it highlights the minimal improvements that would warrant the feasibility of some currently impossible outcome (e.g., achieving a better position in the ranking) or the maximal deteriorations that alternatives can afford to maintain some target result (e.g., not losing their advantage over some other options). The use of a focus group with both experts and participants in the decision making process provided insights on how PFA can support: (i) the creation of arguments in favour or against the respective options under analysis, (ii) understanding of the results' sensitivity with respect to possible changes in the alternatives' performances, (iii) a better informed discussion about the results among the participants in the process, and (iv) the development of new/better alternatives.

Key words: Multiple criteria analysis, Post Factum Analysis, Sensitivity analysis, Urban regeneration, Participation, Greenways

1. Introduction

The analysis of real-world problems requires consideration of multiple conflicting points of view. Nowadays, most decision situations involve economic, environmental, and social considerations to take into account the range of consequences characterising each alternative or course of action [69]. One of the key challenges of decision making problems is thus to cope with the conflicting nature of multiple criteria while taking into account the preferences of Decision Makers (DMs) and stakeholders. Within this context, Multiple Criteria Decision Aiding (MCDA) provides a set of tools and techniques that can support such complex decision making processes [22].

In particular, MCDA methods incorporate the procedures for building a model of the DM's preference on different criteria as well as the algorithms for exploiting an overall preference structure. However, the recommendation that is

*Corresponding author: tel. +48-61 665 3022; fax: +48-61 8771 525.
Email addresses: k.ciomek@gmail.com (Krzysztof Ciomek), V.Ferretti@lse.ac.uk (Valentina Ferretti), milosz.kadzinski@cs.put.poznan.pl (Miłosz Kadziński)

derived from such an exploitation is not equivalent to a decision [4]. Instead, it should be treated as a set of results that have been constructed based on some working assumptions concerning the problem definition and the DM's preferences. The conclusions derived from a specific method are thus often treated as a starting point of a new process of further analysis, interpretation, exploration, and debate [55].

Such a use of the recommendation obtained from MCDA methods is important in view of the vagueness and indeterminacy under which most decisions are made [65]. In this context, Sensitivity Analysis (SA) allows addressing the difficulties in setting the parameter values related to the formulation of a DM's preference model. In particular, SA is used to characterize the impact that changes of the model's parameters (i.e. the DM's preferences) can have on the recommendation, thus allowing to judge whether the latter is sufficiently stable or not. As a result, SA provides the ranges of variation of the model parameters that do not change the recommendation, highlighting as well those critical parameters which, even if just slightly changed, may cause a significant variation in the recommendation [62].

In the vast majority of real-world applications, SA focuses on the impacts that changes in the model parameters, such as criteria weights, can have on the final recommendation (e.g., [17, 65]). Moreover, SA is usually conducted in a "passive" mode, meaning that it identifies modifications of parameters that would lead to different results, but does not actively identify modifications that would allow achieving specific targets.

However, a few approaches instead focus on changes of the performances that are assigned to each action on different criteria, rather than on the parameters of the preference model (for a review, see [36]). The justification for these approaches is that such evaluations can hardly ever be viewed as deterministic entities [65]. Moreover, since these performances represent the current or estimated quality of actions, they are indeed subject to the revisions linked to changes in the impacts associated to the different actions. Obviously, these revisions would affect the final recommendation. This fact has been considered in SA by offering to the DM the possibility to actively formulate the targets that should be achieved by an action with the revised performances.

In this perspective, Wolters and Mareschal [66] and Triantaphyllou and Sánchez [62] have proposed tools to verify how much a performance on a single criterion needs to be improved to make the respective action ranked at the top or above another action, respectively. These ideas have been extended to consider the performance changes on a few criteria simultaneously by minimizing different distance metrics between the original and the changed performance values [31, 32], as well as to take into account more diverse targets (e.g., reaching any higher rank rather than only the first one [3]). The practical usefulness of these methods has been illustrated in view of real-world problems concerning engineering (e.g., [3]), environmental management (e.g., [31]), or investment planning in view of modifying costs (e.g., [54, 66]), location (e.g., [52]), or social aspects (e.g., [54]).

All above studies use a preference model with precise parameter values (i.e., a single pre-defined instance of such a model) to derive the recommendation with the original and modified performance vectors. Moreover, they focus solely on the ranking problems. Finally, they consider only changes required to reach a better result than the current one, and neglect those scenarios in which some outcomes have been already achieved.

To overcome these limitations, [36] proposed a more general framework of Post Factum Analysis (PFA), which addresses multivariate robustness and sensitivity concerns. The most important innovations proposed in this framework can be summarised as follows:

- consideration of the plurality of preference model instances (e.g., multiple additive value functions) compatible with the DM's indirect or imprecise preference information, in accordance with recent MCDA trends (in this case, each target can be achieved in the possible or necessary sense, i.e., respectively, for at least one or all compatible preference model instances [11, 23, 27, 28]);

- focus on a particular aim of the specific decision problem (e.g., the targets that are relevant for multiple criteria sorting [13, 40, 68] are different than those accounted in ranking problems);

- consideration of deteriorations that can be allowed to maintain some results achieved with the current performance vector; this offers a different perspective from the one focused on the improvements needed to achieve the target.

A different framework for SA that also addresses the last concern has been proposed in [53]. It determines the maximal radial performance modification on a single criterion that – when applied to the performances of all actions – allows maintaining the truth of a preference relation established with a single piece-wise linear value function for some pair(s) of actions. Nonetheless, PFA is more general as it considers more than a single value function and performance modifications on multiple criteria at a time, admitting as well more diverse targets. Moreover, when considering the maintenance of some outcome, PFA allows formulating a target which is less demanding than the reference one (e.g., for an action ranked first with the current performance vector one may wonder what deterioration of its performances would allow maintaining a position in top three).

Despite large innovations proposed in PFA [36], the framework still has some limitations which negatively affect its usability. We address these drawbacks in this paper. Firstly, we formalize a problem that needs to be solved in PFA as a Multiple Objective Optimization (MOO) problem in which the performance changes on all individual criteria are simultaneously minimized so that to allow the action with modified performances achieving the target. Such a unified formulation of the problem can be considered irrespective of whether the performance target is already achieved or it is yet to be attained. Precisely, when the deterioration is allowed, the computed minimal performance changes will be negative (thus, indicating the maximal admissible deterioration), whereas in case an improvement is required, the determined minimal changes will be positive.

Furthermore, we make PFA more usable by increasing the variability of accounted criteria scales. Originally, PFA involved radial performance changes, thus multiplying the current evaluations on all criteria by the same factor. Such an approach is too restrictive in the context of MCDA, as it makes PFA applicable only with ratio performance scales. In this paper, we focus on absolute changes rather than relative ones, which implies that PFA can also be used with interval scales. We also demonstrate how the framework can be adapted to deal with ordinal criteria.

Finally, PFA – as originally proposed in [36] – does not allow discriminating the changes that are required/allowed on different criteria, meaning that it requires that the performances on all considered criteria are modified in the same way (e.g., that they are all improved by at least 10% or that they are deteriorated by at most 20%). By incorporating MOO into PFA, we are able to construct all possible scenarios of required or allowed modifications.

The main aim of this paper is to assess the usefulness of the developed PFA framework to support real world decisions. To achieve this aim, we applied PFA on a real project dealing with disused railways requalification in

Italy. We thus used a participatory action research approach (e.g., [48, 50]) within which the second author of the paper assumed the role of a facilitator in the decision making process. The project was thus used as a case study for illustration purposes (i.e. the third important use for case study research according to Siggelkow [58]).

Starting from the recommendation obtained from a previous study involving representatives of a public entity and a private organization [18], we use the individual preferences of the two experts, an arbitrarily selected consensus model and a group decision setting exploiting the space of all possible consensus model instances as input for PFA. We then discuss a variety of results that can be obtained with PFA. Firstly, we conduct a thorough analysis of the performance improvements needed for the two best ranked actions so that each of them could be unanimously considered as the most preferred by the two experts. Then, we focus on the lower-ranked actions to explore the conditions under which they could be ranked better for at least one or all preference model instances relevant for the analysis. Finally, we demonstrate that PFA can deal not only with rank-oriented targets, but it can be also used to compare actions pair-wise, e.g., to determine the performance changes under which a preference relation for some pair of actions would become true.

The results of PFA presented in this paper highlight the required or allowed changes on the individual criteria under consideration, as well as on their subsets relevant for the experts. The latter is particularly interesting as the study of possible scenarios to achieve or maintain some targets reveals that changes on one criterion can be compensated with adequate modifications on another criterion.

Overall, the study proves that PFA is useful for planning and formulating guidelines, but also for verifying the robustness of the recommendation achieved with the current performances. The discussion of the results is enriched with the feedback provided by the two experts relevant for the study during a final focus group.

2. Multiple criteria ranking with an additive value model

This section explains the multiple criteria ranking method that has been used to derive a recommendation within the case study concerning requalification of disused railways. In particular, we define the model that has been employed to represent preferences of different DMs as well as various types of results that have been computed to illustrate the spaces of consensus and disagreement between the experts.

We consider a problem involving a finite set of n actions $A = \{a_1, \ldots, a_i, \ldots, a_n\}$. Each action is evaluated in terms of a set of m criteria $G = \{g_1, \ldots, g_j, \ldots, g_m\}$. Let $g_j(a_i)$ denote a deterministic performance of $a_i \in A$ on $g_j \in G$. To model the DMs' preferences, we use an additive value function [42]:

$$U(a_i) = \sum_{j=1}^{m} w_j \cdot u_j(g_j(a_i)), \tag{1}$$

where u_j is the marginal value function for g_j, $u_j(a_i) \in [0, 1]$ for all $a_i \in A$ and $j = 1, \ldots, m$, and w_j is a relative and non-negative weight associated with g_j. Note that for simplicity of notation, one can write $u_j(a_i)$ instead of $u_j(g_j(a_i))$, for $j = 1, \ldots, m$. The marginal value functions u_j, $j = 1, \ldots, m$, are assumed to be monotonic so that the better the performance $g_j(a_i)$, the greater the marginal value $u_j(a_i)$ that is assigned to it. In particular, for the gain-type criteria $g_j(a_i) \geq g_j(a_k)$ implies $u_j(a_i) \geq u_j(a_k)$. Note that weight w_j can be interpreted as the maximal share that a performance on $g_j(a_i)$ can have in the comprehensive value $U(a_i)$.

For the purpose of interpretability, we assume that comprehensive values $U(a_i)$, $a_i \in A$, are normalized within the range $[0, 1]$. This can be attained by introducing the following constraint:

$$\sum_{j=1}^{m} w_j = 1. \tag{2}$$

The main motivations for selecting an additive value function for the purpose of the case study are linked to the following characteristics: (i) an assumption of compensability among the considered criteria, (ii) the possibility of using piece-wise linear or general marginal value functions (as done in [18]) in view of non-linear preferences of the involved experts for the considered criteria, and (iii) the easiness of interpretation of the numerical scores obtained with this type of model [26]. Additive aggregation implies that a low value on one attribute can be compensated by large values on other attributes. Therefore, this aggregation technique must fulfill relatively strong independence conditions [42] which should be verified in each case. The key condition for the additive form in (1) is a mutual preference independence. Criteria g_j and g_l are preference independent if trade-offs (substitution rates) between g_j and g_l are independent from all other criteria. Mutual preference independence requires that such a requirement is satisfied for all pairs g_j and g_l in G. These independence conditions were tested in the original application as explained in [18].

Using additive value function U requires specification of its parameters. These can be elicited either directly or indirectly [23, 34]. According to the direct elicitation approach, the DM needs to specify values related to the formulation of marginal value functions and weights. Conversely, in the indirect approach, these values are inferred from some holistic decision examples (e.g., a complete ranking of a subset of reference actions [33] or some incomplete pair-wise comparisons [27]). Recent experimental studies [10, 37] indicate that an indirect elicitation of preferences is more suitable when dealing with problems that involve numerous alternatives, few criteria and few parameters of the assumed preference model to be elicited. Since none of these conditions was satisfied in the presented case study, other techniques have been used for the elicitation of marginal value functions and weights (see [18] and Section 3). Whichever the type of preference information, let us denote all of its pieces (e.g., weight constraints or holistic pair-wise comparisons) provided by the DMs by S.

When S precisely defines the parameters of U, there exists a single value function compatible with the DMs' preferences. Its application on A leads to assigning a comprehensive value $U(a_i)$ to each action $a_i \in A$, which allows establishing a complete pre-order on A. In this way, all pairs of actions are made comparable. Precisely, if $U(a_i) > U(a_k)$, then a_i is preferred to a_k, whereas if $U(a_i) = U(a_k)$, then a_i and a_k are considered indifferent. Consequently, the rank of action a_i is defined as follows [39]:

$$rank(U, a_i) = 1 + \sum_{a_k \in A \setminus \{a_i\}} h(U, a_i, a_k), \text{ where} \tag{3}$$

$$h(U, a_i, a_k) = \begin{cases} 1, & \text{if } U(a_k) > U(a_i), \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

For example, if there are two actions $a_k \in A \setminus \{a_i\}$ with comprehensive values $U(a_k)$ better (greater) than $U(a_i)$, a_i would be ranked third.

On the contrary, if S leaves some freedom with respect to the possible parameter values related to the formulation of marginal functions u_j and/or criteria weights w_j, $j = 1, \ldots, m$, there exists a set $\mathcal{U}(S)$ of additive value models compatible with the DMs' preferences S. In this case, one may either arbitrarily select a single representative model from $\mathcal{U}(S)$, or account for the robustness concerns when working out a recommendation by considering all compatible models in $\mathcal{U}(S)$ simultaneously.

When using a single value function for deriving a recommendation one may assign a score to each action and easily assess the weights of criteria. In this sense, its use for decision aiding may be deemed less abstract than that of the whole set of functions [24]. In fact, the latter exhibits the impact of the plurality of compatible preference models on the recommendation, though without presenting these (possibly infinitely many) models to the DMs. Among the selection procedures of a representative value function, the most widely used techniques consist in selecting a central [5, 15, 56] or a mean [1, 33, 35] model. The former corresponds to a center of the polyhedron representing all compatible models, whereas the latter is defined as an average from all models or their subset.

To avoid arbitrarily selecting a single representative model from $\mathcal{U}(S)$, one may perform robustness analysis, thus, implementing the prudence principle in decision aiding. In general, robustness analysis is related to the examination of input variability, imprecision, and uncertainty on the stability of the proposed recommendation [39, 59, 64]. In our case, these are related to the existence of multiple additive value functions in $\mathcal{U}(S)$ which are relevant for the analysis. Obviously, each compatible value function $U \in \mathcal{U}(S)$ may lead to a different ranking. In this perspective, a conclusion is considered to be robust if it is valid for all or most acceptable values for the model parameters.

When dealing with ranking problems, the robustness of the suggested recommendation may refer to the stability of pair-wise preference relation or ranks achieved by the actions. When comparing pairs of actions in A, the following results can be derived:

- the necessary preference relation $\succsim^N_{\mathcal{U}(S)}$ that holds for a pair (a_i, a_k) iff a_i is preferred to a_k for all value functions in $\mathcal{U}(S)$ [27];
- the possible preference relation $\succsim^P_{\mathcal{U}(S)}$ that holds for a pair (a_i, a_k) iff a_i is preferred to a_k for at least one value function in $\mathcal{U}(S)$ [27];
- pair-wise winning index $PWI(a_i, a_k)$ that indicates the share of value functions in $\mathcal{U}(S)$ for which a_i is preferred to a_k [39, 46].

The relevant robust rank-related results in terms of investigating the consequences of applying $\mathcal{U}(S)$ on A are the following:

- the highest (best) $P^*_{\mathcal{U}(S)}(a_i)$ and the lowest (worst) $P_{*,\mathcal{U}(S)}(a_i)$ ranks achieved by a_i in $\mathcal{U}(S)$ [38];
- rank acceptability index $RAI(a_i, r)$ that indicates the share of value functions in $\mathcal{U}(S)$ for which a_i achieves r-th rank [16, 39, 44]; defined in this way, RAI quantifies how probable it is for an action to be ranked in a given position.

The necessary, possible, and extreme results are derived by solving some dedicated Linear Programming (LP) models [27, 38], whereas the values of acceptability indices are estimated using Monte Carlo simulation [60]. For this purpose, we use the Hit-And-Run algorithm [61] which has been designed to appropriately sample the space of uniformly distributed value functions $\mathcal{U}(S)$.

3. The context of the case study

This study builds on and significantly extends the results obtained during the decision aiding process developed in [18]. The study presented in [18] aimed at supporting the local authority in the Piedmont Region in understanding what could be the most suitable requalification project for a recently abandoned railway line. The present study starts from the recommendation obtained from the above mentioned process, and focuses on the preferences of the two most important stakeholders, thus further developing the analysis through the PFA framework [36].

From the territorial context point of view, it is interesting to highlight that in Italy there are more than 7500km of abandoned railways, 50% of which have been evaluated as suitable to be recovered for touristic purposes and ecological valorization. In 2012, due to cost-saving measures, the Piedomont Region (North West of Italy) decided to dismiss twelve passenger railway lines, characterised by low patronage, and replaced them by bus services [18]. In this paper, we focus on the abandoned railway line *Pinerolo-Torre Pellice*, which has emerged as the most strategic one to be recovered first [18]. We consider the following five actions for the requalification of this railway:

- greenway (a_1), i.e., the conversion of the 16.5km of abandoned railway into a green corridor linking five municipalities in a mixed rural and urban area;
- rail-banking (a_2), i.e., ordinary maintenance works on the railway tracks in order to ensure the standards of quality, security and efficiency that are compatible with a possible reopening of the railway tracks in the future;
- extension of the urban railway service (a_3), i.e., an extension of line 2 of the urban railway service (which has been created with the aim of improving the efficiency of the connections between Turin, i.e., the capital of the region, and the more peripheral cities) in order to include the municipalities that were crossed by the railway;
- old station recovery (a_4), i.e., a recovery of the old station building in the Municipality of Luserna San Giovanni, which has been estimated to be the most strategic one to be recovered for touristic and recreational purposes;
- no action (status quo, a_5), i.e., not taking any action and letting the abandoned railway exposed to natural degradation, structural failures and to the risk of being used as illegal landfill.

The requalification of an abandoned railway line is a complex decision making problem, involving multiple and conflicting perspectives. Consequently, to identify which is the best action for the qualification of the abandoned railway line, a set of nine criteria has been identified for the evaluation of the actions [18]. The reader can refer to the description of the original intervention in [18] for the detailed explanation of all criteria. In particular, semi-structured interviews with the experts were used to support the definition of a comprehensive set of both stakeholders to be involved in the collaborative process and objectives to be achieved with the proposed requalification strategy [18].

Further insights for the identification of the impacts came from the analysis of the scientific literature as well as from the legislative requirements in the field of sustainability assessments of the territorial transformation processes. The identified criteria have been grouped as follows:

- environmental factors: creation of new green areas (g_1; in m^2), compatibility with the present land use (g_2; ordinal scale), duration of the construction works (g_3; in months), and landscape impacts (g_4; ordinal scale);
- socio-economic factors: construction costs (g_5; in euro), new jobs (g_6; quantitative scale), impacts on the touristic sector (g_7; ordinal scale), potential users (g_8; cardinal scale), and presence of attractions (g_9; ordinal scale).

It is worth highlighting that the *duration of the construction works* is included among the environmental factors, because it has been defined as the *time needed for the realization of each alternative project, thus working as a proxy attribute for the interferences with the natural ecosystems in the surrounding areas which could be affected by the construction works* (see page 39 in [18]). Table 1 summarizes the performances of each action, showing as well the direction of preference on each criterion, the performance scales and the ranges (minimum and maximum for each criterion). Performances on the ordinal scales (O) are encoded with integers preserving the preference order.

Table 1: The actions, criteria and performances for the problem under analysis (↑ and ↓ indicate whether a criterion has to be maximized or minimized, respectively; O and I indicate ordinal and interval scales, respectively, source of the performances: [18]).

Alt.	$g_1 \uparrow$	$g_2 \uparrow$	$g_3 \downarrow$	$g_4 \uparrow$	$g_5 \downarrow$	$g_6 \uparrow$	$g_7 \uparrow$	$g_8 \uparrow$	$g_9 \uparrow$
a_1	165000	Good (1)	12	V. positive (3)	830000	4	High (2)	75000	78
a_2	0	V. good (2)	1	Irrelevant (1)	170000	0	None (0)	0	0
a_3	0	V. good (2)	1	Irrelevant (1)	170000	3	Med. (1)	249200	33
a_4	40000	Low (0)	5	V. positive (3)	240000	5	High (2)	19400	32
a_5	0	V. good (2)	0	Negative (0)	0	0	None (0)	0	0
Scale	I	O	I	I	I	I	O	I	I
Range	$[0, 165000]$	$\{0, 1, 2\}$	$[0, 12]$	$\{0, 1, 2, 3\}$	$[0, 830000]$	$[0, 5]$	$\{0, 1, 2\}$	$[0, 249200]$	$[0, 78]$

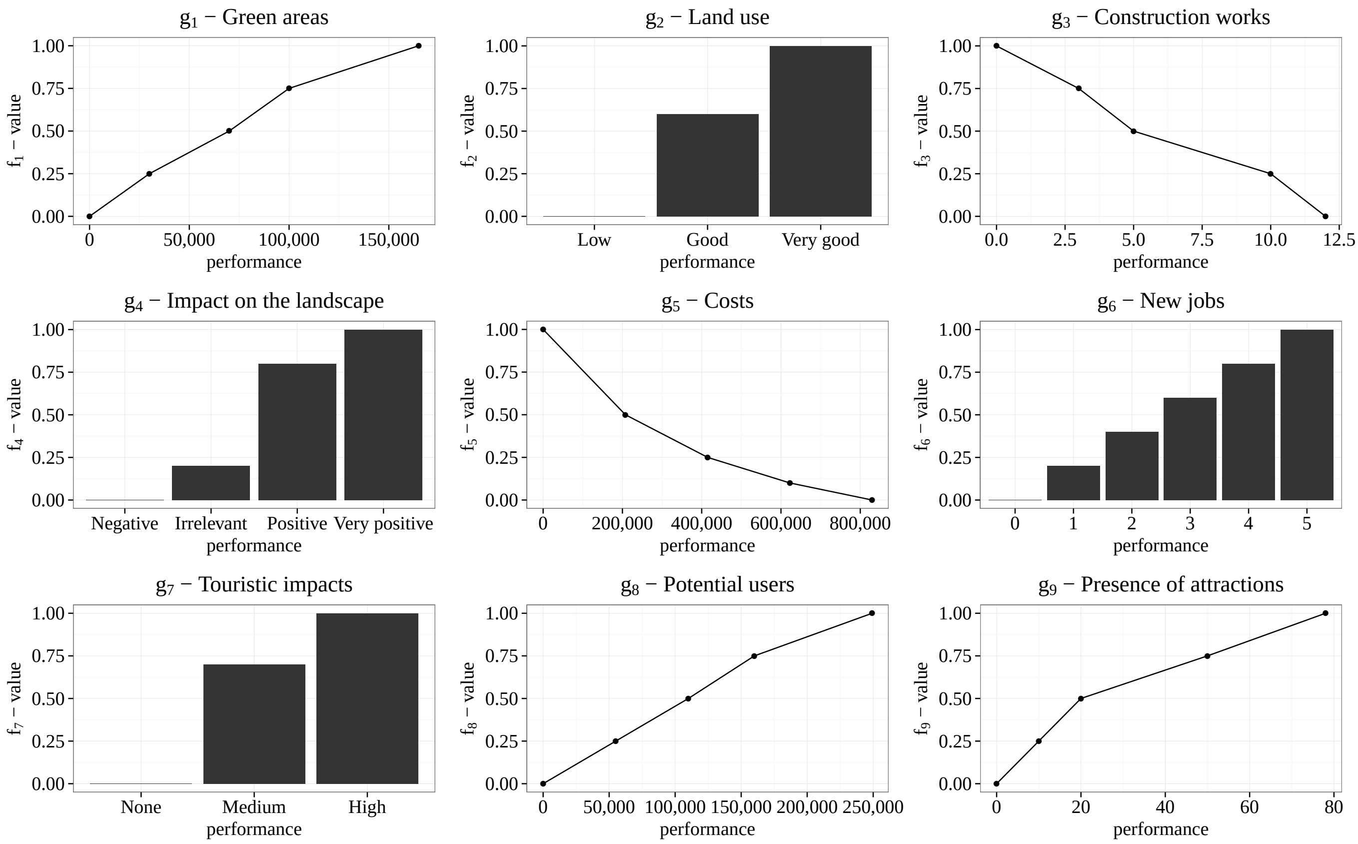


Figure 1: Marginal value functions elicited for the decision making problem under analysis.

Figure 1 shows the marginal value functions elicited for each criterion. The interested reader can refer to [18] for a detailed account of the value functions elicitation process. For the purpose of this new study, it is enough to highlight that for the quantitative criteria the bisection elicitation protocol has been used, while for the qualitative criteria the direct elicitation protocol has been applied. As a result of the value function elicitation process, the standardized values of the actions are presented in Table 2. In particular, this process allowed to have all criteria expressed on a common scale, i.e. from 0 (poor performance and low objective achievement) to 1 (good performance and full objective achievement), thus making them comparable and suitable for aggregation.

Table 2: Marginal values for the considered actions.

	Action	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9
a_1	Greenway	1.00	0.60	0.00	1.00	0.00	0.80	1.00	0.34	1.00
a_2	Rail-banking	0.00	1.00	0.92	0.20	0.59	0.00	0.00	0.00	0.00
a_3	Transport	0.00	1.00	0.92	0.20	0.59	0.60	0.70	1.00	0.61
a_4	Old station	0.31	0.00	0.50	1.00	0.46	1.00	1.00	0.09	0.60
a_5	No action	0.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00

While the original decision support process involved six participants (experts and stakeholders) [18], in this study we focus only on the two most important stakeholders and on their respective preferences. These are the Transportation Authority (private organisation) and the Piedmont Region Authority (public entity). The reason why we selected these two stakeholders is linked to the results obtained by [18]. In particular, the two best performing actions obtained in the final ranking of the original study are the creation of a greenway (first position in the final ranking) and the extension of the urban transportation system (second position in the final ranking and very close to the first one). These actions are indeed the two favoured options, the greenway being supported by the Piedmont Region Authority and the extension of the urban railway system being supported by the Transportation Authority.

In the remainder of this paper, DM_1 represents the consultant for the Transportation Authority who was involved as an expert in the field of transportation engineering, while DM_2 represents the consultant for the Regional Authority who was involved as an expert in the field of landscape ecology. The weights elicited from these experts using the swing-weights procedure [2] are presented in Table 3 (for details on the use of the procedure, see [18]). These do differ as, e.g., DM_1 assigns greater weight to land use compatibility (g_2), landscape impacts (g_5), construction costs (g_5), and the number of potential users (g_8), whereas DM_2 is more concerned about green areas (g_1), duration of the construction works (g_3), touristic impacts (g_7), and presence of attractions (g_9).

Table 3: Weights elicited from the two experts: DM_1 and DM_2. AVG represents an average between the two estimates.

g_i	Criterion	DM_1	DM_2	AVG
g_1	Green areas	0.117	0.172	0.145
g_2	Land use	0.133	0.086	0.110
g_3	Construction works	0.083	0.138	0.111
g_4	Impact on the landscape	0.167	0.103	0.135
g_5	Costs	0.160	0.094	0.127
g_6	New jobs	0.020	0.047	0.033
g_7	Touristic impacts	0.080	0.156	0.118
g_8	Potential users	0.200	0.125	0.162
g_9	Presence of attractions	0.040	0.078	0.059

We treat the weights inferred for the individual DMs as bounds of the weight space that is relevant for the problem. We will denote this space by $\mathcal{U}$. Note that each feasible weight vector corresponds to a compatible additive value function. To this end, in Table 3 we also present the average weight vector. Let us denote the value functions obtained by combining the marginal value functions with the respective weights by U_{DM_1}, U_{DM_2}, and U_{AVG}. Obviously, $U_{DM_1}, U_{DM_2}, U_{AVG} \in \mathcal{U}$.

The rankings obtained with the weights provided by DM_1 and DM_2, as well as for the average vector of weights are presented in Table 4. On one hand, although differing in terms of the comprehensive values, all rankings agree with respect to the bottom three actions (a_4, a_5, and a_2). On the other hand, for DM_1 the extension of the urban railway service (a_3) is ranked first, while for DM_2 and for the average model the greenway (a_1) is the most preferred option. We highlight that a_1 derives its advantage from g_1, g_4, g_7, and g_9, whereas the most favourable features of a_3 include g_2, g_3, g_7, and g_8.

Table 4: Comprehensive values and corresponding rankings for the two experts and for the average model.

	a_1	a_2	a_3	a_4	a_5	Ranking
DM_1	0.568	0.338	0.630	0.460	0.377	$a_3 \succ a_1 \succ a_4 \succ a_5 \succ a_2$
DM_2	0.642	0.289	0.599	0.531	0.318	$a_1 \succ a_3 \succ a_4 \succ a_5 \succ a_2$
AVG	0.605	0.313	0.614	0.495	0.347	$a_3 \succ a_1 \succ a_4 \succ a_5 \succ a_2$

To investigate the stability of the provided recommendation within the space of possible consensus weight vectors $\mathcal{U}$ for the two DMs, we performed a robustness analysis. In Table 5, we present the extreme ranks, i.e., the highest and lowest ranks achieved by each action for some weight vector relevant for the study [38]. Their analysis indicates that:

- the only potential best options are a_1 ad a_3 (their best ranks are equal to one; $P^*_{\mathcal{U}}(a_1) = P^*_{\mathcal{U}}(a_3) = 1$), which confirms the outcomes derived from the analysis performed for the DMs' individual weights;
- when considering the potentially optimal actions, the relative performance of a_1 is more stable than that of a_3 as its rank may drop only to the second place in the worst case ($P_{*,\mathcal{U}}(a_1) = 2 < P_{*,\mathcal{U}}(a_3) = 3$);
- a_3 and a_4 are the most sensitive to the selection of a single compromise weight vector in $\mathcal{U}$ as they may achieve positions in the range $[1, 3]$ or $[2, 4]$, respectively;
- the bottom position of rail-banking (a_2) is stable across the set of feasible weight vectors $\mathcal{U}$.

Table 5: Positions achieved by the actions in the individual rankings and extreme ranks achieved within the space of possible consensus weight vectors.

Preference model	a_1	a_2	a_3	a_4	a_5
U_{DM_1}	2	5	1	3	4
U_{DM_2}	1	5	2	3	4
U_{AVG}	2	5	1	3	4
$\mathcal{U}$ ($[P^*_{\mathcal{U}}(a_i), P_{*,\mathcal{U}}(a_i)]$)	$[1, 2]$	$[5, 5]$	$[1, 3]$	$[2, 4]$	$[3, 4]$

The investigation of the stability of the recommendation for pairs of actions is represented by the necessary preference relation [27] (see Figure 2). The truth of the necessary relation for a pair of actions indicates that one of them is preferred to another for all weight vectors relevant for the analysis. Thus, such information needs to be regarded with certainty. For example, a_1 is ranked better than a_2, a_4, and a_5 in the entire space of possible consensus weight vectors $\mathcal{U}$.

Conversely, there are three pairs of actions: (a_1, a_3), (a_3, a_4) and (a_4, a_5), which are incomparable in terms of the necessary relation. These are not connected by an arc in Figure 2. This means that for some feasible weight vectors one action is ranked better than the other, whereas for some other admissible weight vectors, they are ranked in an inverse order. Consequently, there is an ambiguity with respect to the identification of the more favourable action in such pairs.

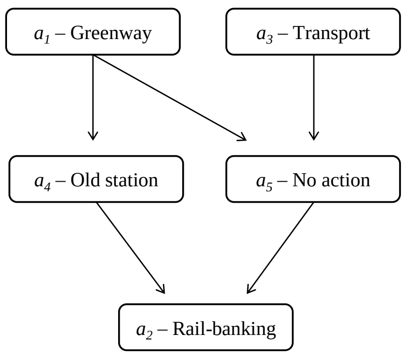


Figure 2: The necessary preference relation.

To enrich the conclusions that can be derived from the analysis of the necessary and extreme results, one can refer to the stochastic acceptability indices [39, 60] (see Table 6). These have been derived from Monte Carlo simulation based on 10000 samples of the admissible weight vectors [61]. Each acceptability index can be interpreted as the probability of observing the underlying part of the recommendation in the set of feasible weight vectors. For example, the Rank Acceptability Indices (RAIs) for a_3 indicate that for about 61% of possible consensus weight vectors in $\mathcal{U}$, a_3 is ranked first ($RAI(a_3, 1) = 0.61$), for about 39% of feasible weight vectors it is ranked second ($RAI(a_3, 2) = 0.39$), while the estimates of the probability of achieving any rank outside the top two are equal to zero. In particular, although extreme ranking analysis indicates that a_3 can be possibly ranked third, $RAI(a_3, 3)$ is zero, which means that the probability of such an outcome is negligible.

Table 6: Stochastic acceptability indices.

(a) Pair-wise winning indices PWIs

	a_1	a_2	a_3	a_4	a_5
a_1	0.00	1.00	0.39	1.00	1.00
a_2	0.00	0.00	0.00	0.00	0.00
a_3	0.61	1.00	0.00	1.00	1.00
a_4	0.00	1.00	0.00	0.00	1.00
a_5	0.00	1.00	0.00	0.00	0.00

(b) Rank acceptability indices RAIs

	1	2	3	4	5
a_1	0.39	0.61	0.00	0.00	0.00
a_2	0.00	0.00	0.00	0.00	1.00
a_3	0.61	0.39	0.00	0.00	0.00
a_4	0.00	0.00	1.00	0.00	0.00
a_5	0.00	0.00	0.00	1.00	0.00

The analysis of stochastic acceptability indices allows drawing the following conclusions that cannot be derived when considering the necessary and extreme outcomes:

- when it comes to the potentially top actions, a_3 compares positively to a_1 being ranked better for 61% of the relevant models (see Table 6a) and b));
- the preference of a_5 over a_4 and a_4 over a_3, although possible, is extremely unlikely; indeed, the estimated values of $PWI(a_5, a_4)$ and $PWI(a_4, a_3)$ are equal to zero (see Table 6a);
- the rank acceptabilities confirm a clear status of a_4 and a_5 as the third and fourth most preferred actions, respectively (see Table 6b).

The above analyses and graphs were presented at a final focus group carried out together with the two experts representing the perspective of DM_1 and DM_2 from the study developed in [18]. In particular, the second author of this paper worked as a facilitator during the focus group, explaining the results and taking notes about the feedback and comments provided by the participants. Overall, the participants in the focus group noted that the outcomes of robustness analysis are useful for:

- screening the considered actions to keep only the best ones (a_1 and a_3) for a more detailed analysis (e.g., technical feasibility studies);
- gaining valid arguments in favour of one option over another as derived, e.g., from the comparison of PWIs for a_3 and a_1;
- identifying some reasonably good actions (e.g., a_4 which is ranked second in the best case) that can be implemented jointly with some other option (e.g., with a_1);

- constructing arguments toward the achievement of a consensus, which is supported by the graphical and tabular presentation of the results.

4. Performance-based Post Factum Analysis

The analysis of the recommendations may stimulate further questions from the DMs. Indeed, knowing the recommended order of actions or the preference relation imposed on the set of actions, the experts may wonder [36]:

- "what improvement of the performances of a given action should be made so that it can achieve a better position or become preferred to another action?";
- "what deterioration of performances it can afford to maintain some position or to be still considered more favourable than some other actions?".

Thus, when analyzing the performance of action $a_i \in A$ in view of multiple criteria ranking, one may consider two types of targets:

- a_i is preferred to a_k, denoted by $a_i \succsim a_k$;
- a_i achieves r-th rank, denoted by $a_i \to r$.

They concern preference- and rank-related perspectives, respectively. Hence, we can either collate actions "one vs one" or focus on the performance of each individual action, at the same time confronting it with all remaining actions jointly.

Whichever the target, its achievement can be analyzed in view of a single value function U or a set of value functions $\mathcal{U}$. The former is useful, e.g., when considering the recommendation obtained for an individual DM with a precisely elicited or derived preference model. The latter can be applied either for a single DM whose preference model has been elicited in an indirect way (which usually leaves some freedom to the variability of model parameters) or for a set of possible consensus models for a group of DMs. Furthermore, when accounting for multiple value functions in $\mathcal{U}$, the achievement of the target may be considered in the possible or necessary sense. Consequently, we may require that it is achieved for at least one or all value functions, respectively.
We will use the following notation with respect to the targets:

$$T_Z^Y(X), \tag{5}$$

where:

- $X \in \{a_i \succsim a_k,\ a_i \to r\}$ represents the target under consideration;
- $Y \in \{U,\ \mathcal{U}\}$ represents either a single value function U or a set of relevant functions $\mathcal{U}$;
- $Z \in \{P,\ N\}$ with P and N representing, respectively, the possibility or the necessity of achieving the target (note that in case $Y = U$, Z can be neglected).

For example, $T^U(a_i \to r)$ concerns a_i being ranked r-th in the order imposed by U, whereas $T_N^{\mathcal{U}}(a_i \succsim a_k)$ is related to a_i being preferred to a_k for all value functions in $\mathcal{U}$. We denote the achieved target by $T_Z^Y(X) = 1$, and we write $T_Z^Y(X) = 0$, otherwise.

Let us denote the absolute changes of the performance of $a_i \in A$ on g_j, $j = 1, \ldots, m$, by $b(a_i) = [b_1^i, \ldots, b_j^i, \ldots, b_m^i]$. For the ratio and interval scales, b_j^i is a performance difference, whereas for an ordinal scale b_j^i is interpreted as a certain number of ordinal levels. When applying $b(a_i)$ to the performances of a_i, one obtains a new action a_i^b. Clearly, for the criteria with numeric scales $g_j(a_i^b) = g_j(a_i) + b_j^i$, whereas for the criteria that do not allow for a relative degree of difference $g_j(a_i^b)$ is obtained through changing $g_j(a_i)$ by b_j^i levels.

The allowed performance changes are defined within space $B(a_i)$. This space is delimited by two types of constraints. Firstly, on each criterion g_j, the modified performance cannot be outside the criterion's range $[g_j^{min}, g_j^{max}]$ (thus, $g_j^{min} - g_j(a_i) \leq b_j^i \leq g_j^{max} - g_j(a_i)$). Note that, depending on the decision context, these extreme performances can be either pre-defined, thus, delimiting the performance scale, or derived from the performances observed for actions in A. Secondly, one can specify additional limitations on $B(a_i)$ using some domain knowledge. For example, the DM may require that the performance of a_i on g_j may change only by $p\%$ (i.e., $-p\% \cdot g_j(a_i) \leq b_j^i \leq p\% \cdot g_j(a_i)$) when a ratio scale is considered or that the change of p performance levels is admissible in case an ordinal scale is employed (i.e. $-p \leq b_j^i \leq p$). If the performance on a particular criterion cannot be changed, then $b_j^i = 0$.

Furthermore, if a target is yet to be achieved, the action's performances need to be improved so it is natural to assume that $b \geq 0$. Otherwise, if a target is just to be maintained, the deteriorations of performances can be tolerated, and, thus, we assume that $b \leq 0$.

Let us denote the admissible performance changes for $a_i \in A$ and g_j, $j = 1, \ldots, m$, derived from all these constraints by $b_j^{i,min}$ and $b_j^{i,max}$. Overall, a space of possible performance modifications that can be applied to $a_i \in A$ on all criteria can be defined as:

$$B(a_i) = \left\{ b(a_i) : \forall_{j=1\ldots m}\ b_j^{i,min} \leq b_j^i \leq b_j^{i,max} \right\}. \tag{6}$$

Considering a certain target $T_Z^Y(X)$ involving action $a_i \in A$, let $B(T_Z^Y(X))$ denote the set of all vectors $b(a_i) \in B(a_i)$ for which this target is satisfied, i.e., $T_Z^Y(X) = 1$. Obviously, we are interested in the minimal performance changes for which this happens. On one hand, in case the target is not achieved with the current performances, this would allow identifying the minimum improvements that are required to achieve it. On the other hand, when the target is already achieved, one would identify the greatest admissible deterioration of performances for which the target is still maintained. The problem that needs to be solved for this purpose can be formulated as follows:

$$Minimize\ b_1^i, \ldots, b_j^i, \ldots, b_m^i, \tag{7}$$

s.t.:

$$\left.\begin{array}{l}
Y \in \{U, \mathcal{U}\}, \\
\text{ for all } j = 1, \ldots, m: \\
b_j^{i,min} \leq b_j^i \leq b_j^{i,max}, \\
g_j(a_i^b) = g_j(a_i) + b_j^i, \\
\text{if } Z = N, \text{ then for all } U \in \mathcal{U}: \\
\quad \text{if } \; X = a_i \succsim a_k: \\
\quad\; U(a_i^b) = \sum_{j=1}^m w_j \cdot u_j(g_j(a_i^b)) \geq U(a_k) = \sum_{j=1}^m w_j \cdot u_j(g_j(a_k)), \\
\quad \text{if } \; X = a_i \rightarrow r: \\
\quad\; U(a_i^b) = \sum_{j=1}^m w_j \cdot u_j(g_j(a_i^b)) \geq [\sum_{j=1}^m w_j \cdot u_j(g_j(a_k))] - M \cdot v_{a_k}, \\
\qquad \text{for all } a_k \in A \setminus \{a_i\}, \\
\quad \sum_{a_k \in A \setminus \{a_i\}} v_{a_k} \leq r - 1, \\
\quad v_{a_k} \in \{0, 1\}, \text{ for all } a_k \in A \setminus \{a_i\}.
\end{array}\right\} E(T_Z^Y(X), a_i^b)$$

Let us denote the subset of Pareto-optimal solutions for which $T_Z^Y(X) = 1$ with $B^{PF}(T_Z^Y(X)) \subseteq B(T_Z^Y(X))$. For each performance change vector in $B^{PF}(T_Z^Y(X))$ there is no other vector in $B(T_Z^Y(X))$ that would allow achieving the target with not greater changes on all criteria and strictly less change on at least one criterion.

An exemplary interpretation of these subsets for a problem involving a pair of criteria (g_1, g_2) is presented in Figure 3. The space $B(a_i)$ of admissible performance changes (b_1, b_2) that can be applied to $a_i \in A$ is delimited by the dashed lines. In particular, $b_1^{min} < 0$ and $b_1^{max} > 0$ are the extreme allowed modifications of $g_1(a_i)$. A grayed subspace of $B(a_i)$ represents all performance changes $B(T_Z^Y(X))$ that allow achieving a certain target $T_Z^Y(X)$. These involve sufficiently great improvements on both criteria $(b_1, b_2 > 0)$, but the target can be also achieved when a small deterioration on one criterion (e.g., $b_1 < 0$) is compensated with an adequately greater improvement on the other criterion (e.g., $b_2 > 0$). Within $B(T_Z^Y(X))$ a subset of Pareto-optimal solutions $B^{PF}(T_Z^Y(X))$ is represented with a solid line. For each vector of performance modifications in $B^{PF}(T_Z^Y(X))$, there is no other vector that would be more favourable than it, thus, allowing to achieve $T_Z^Y(X)$ with less performance changes.

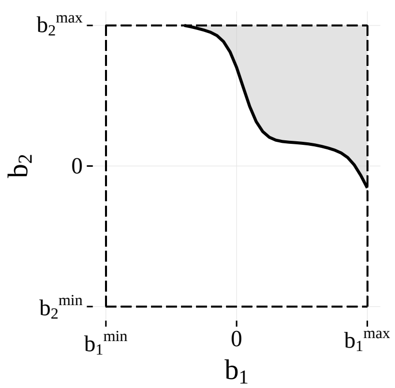


Figure 3: An exemplary two-dimensional space of allowed performance changes for action a_i in view of target $T_Z^Y(X)$: $B(a_i)$ is delimited by dashed lines, $B(T_Z^Y(X))$ is grayed, and $B^{PF}(T_Z^Y(X))$ is represented with a solid black line.

Note that all solutions in $B^{PF}(T_Z^Y(X))$ depend on the set of compatible value functions that are involved in the specification of constraint set $E(T_Z^Y(X), a_i^b)$. When considering some target $T_N^Y(X)$ in view of its achievement in the necessary sense, the obtained solutions need to guarantee that the target is satisfied for all compatible value functions. For this reason, one may claim that the solutions in $B^{PF}(T_N^Y(X))$ do not depend on any particular function. However, it is important to emphasize that for some compatible functions the target may be satisfied even with less performance changes, and it is the least advantageous function in terms of the target achievement that affects the performance modifications contained in the respective solution. In the same spirit, when accounting for target $T_P^Y(X)$ that has to be satisfied in the possible sense (i.e., for at least one compatible value function), each solution in $B^{PF}(T_P^Y(X))$ is associated with some function that can be considered as the most advantageous in terms of the target achievement. These functions can be, in general, different for various solutions contained in $B^{PF}(T_P^Y(X))$.

Whichever target T, the respective set of solutions $B^{PF}(T_Z^Y(X))$ can be exploited to identify a single performance change vector that would be most preferred by the DM. For this purpose, one can apply interactive multiple objective optimization (MOO) methods [6], or optimize the Minkowski distance:

$$\|\hat{b}(a_i)\|_\lambda = \left(\sum_{j=1}^{m} \hat{b}_j(a_i)^\lambda \right)^{\frac{1}{\lambda}} \tag{8}$$

where $\hat{b}_j(a_i)$, $j = 1, \ldots, m$, corresponds to $b_j(a_i)$ normalized in the space of possible performance modifications, i.e., $\hat{b}_j(a_i) = \frac{b_j(a_i) - b_j^{min}}{b_j^{max} - b_j^{min}}$. Clearly, it is also possible to optimize a distance involving the non-normalized performance modifications, i.e., $\|b(a_i)\|_\lambda = \left(\sum_{j=1}^{m} b_j(a_i)^\lambda\right)^{\frac{1}{\lambda}}$, but in this case one needs to be aware of the impact that different criteria scales may have on the result. Note that the Minkowski distance is typically used for $\lambda = 1$, $\lambda = 2$, or $\lambda = \infty$, which correspond to the Manhattan, Euclidean and Chebyshev distances, respectively. Such a distance has been widely used in MCDA, e.g., in the reference point MOO approaches [47] or the TOPSIS method [30].

5. The application of Post Factum Analysis to support the requalification of abandoned railway lines

For the problem concerning the requalification of abandoned railway lines in Piedmont (see Section 3), the experts were interested in the potential improvements and deteriorations of performances concerning different targets. To this end, the following constraints have been identified:

- performances related to *land use* (g_2) cannot be changed (thus, for $a_i \in A$, $b_2^i = 0$);
- action's *touristic impacts* (g_7) can be realistically changed only by one performance level (thus, for $a_i \in A$, $-1 \leq b_7^i \leq 1$);
- on all criteria the modified performance cannot be outside the range delimited by the extreme performances observed for the set of actions (thus, e.g., for $a_i \in A$, $0 \leq g_1(a_i) + b_1^i \leq 165000$ or, equivalently, $b_1^{i,min} = 0 - g_1(a_i)$ and $b_1^{i,max} = 165000 - g_1(a_i)$);
- although some performances of a_5 – *no action* – can be changed (e.g., its performance on g_4 may be improved by creating new natural landmarks), these modifications represent remote scenarios that are out of control for

the DMs; thus, we assume for $j = 1, \ldots, m$, $b_j^5 = 0$.

Furthermore, the experts defined the subsets of criteria under interest when conducting PFA for a particular target. For cognitive limitations, these subsets were limited to triplets of criteria for which the performances can be simultaneously changed.

5.1. *Optimization algorithms used in the case study*

Since problem (7) is a non-linear multiple objective optimization problem, it cannot be solved effectively with the contemporary solvers [6, 36]. However, if one considers a performance change b_j^i of $a_i \in A$ on just a single criterion g_j, $j = 1, \ldots, m$, the problem can be approached with simple optimization techniques. In particular, we applied a binary search (bisection) method [43], which repeatedly bisects the interval of admissible performance changes $[b_j^{i,min}, b_j^{i,max}]$, and then selects for further processing a subinterval in which a minimal change must lie.

Algorithm 1 presents a pseudo-code for the binary search method adapted to compute the minimal performance modification when dealing with qualitative criteria scales. To verify if the target is already achieved with the currently tested performance modification $b_j^i(T_Z^Y(X))$, we check if $E(T_Z^Y(X), a_i^b)$ is feasible. If so, the proposed change is sufficient to achieve $T_Z^Y(X)$. Thus, we modify the upper bound of the exploited interval $[b_j^{i,down}, b_j^{i,up}]$ of admissible changes to further investigate if the target can be achieved with even less modification of performances. Otherwise, the proposed change is too small and we modify the lower interval bound. The procedure is repeated until a width of the interval $[b_j^{i,down}, b_j^{i,up}]$ in which the optimal solution must lie is greater than the required precision γ of the identified solution. In our study, we applied $\gamma = 10^{-15}$.

Algorithm 1 The binary search method for computing the minimal performance change $b_j^i(T_Z^Y(X))$ for the qualitative scales that allows a_i to achieve $T_Z^Y(X)$.

Require: $a_i \in A$, an action which should achieve $T_Z^Y(X)$ by changing its performance on g_j.
Require: $b_j^{i,min}$ and $b_j^{i,max}$, the extreme allowed performance changes.
Require: γ, the precision of the computed result (e.g., $\gamma = 0.00001$).
Ensure: $b_j^i(T_Z^Y(X))$, the minimal performance change needed to achieve $T_Z^Y(X)$.

$b_j^{i,down} = b_j^{i,min}$ and $b_j^{i,up} = b_j^{i,max}$.
while $b_j^{i,up} - b_j^{i,down} \geq \gamma$ **do**
    $b_j^i(T_Z^Y(X)) = (b_j^{i,up} - b_j^{i,down})/2$.
    if $E(T_Z^Y(X), a_i^b)$ is feasible **then**
        $b_j^{i,up} = b_j^i(T_Z^Y(X))$.
    else
        $b_j^{i,down} = b_j^i(T_Z^Y(X))$.
    end if
end while

In case the performances on at least two criteria are to be changed, to identify a reliable approximation of $B^{PF}(T_Z^Y(X))$, one needs to apply more advanced optimization techniques. For this purpose, we used an evolutionary algorithm called NSGA-II [14]. Its role is to estimate meta-heuristically the Pareto front $B^{PF}(T_Z^Y(X))$. NSGA-II starts with an initialization of a random parent population of solutions of size N. In our case, each solution represents a vector of performance changes. Then, the offspring of the same size is created using the usual genetic operators. The parents and their offspring are combined to obtain a population of size $2N$, which is sorted using a fast non-dominated sorting algorithm and a crowding distance [14]. The new population of size N is constructed so that to promote solutions which are non-dominated or dominated by as few other solutions as possible and to maintain a uniform spread-out of the approximated Pareto front (for details, see [14, 41]). In this way, the evolutionary process allows generating the population of solutions which are more favourable (thus, representing less performance modifications which are required to achieve the target under consideration) than those contained in the preceding populations. The process is iterated until a stopping criterion is met.

For the purpose of our study, we suitably adapted the implementation of NSGA-II available in [49]. Moreover, for the 2- and 3-criteria analysis we used a population consisting of, respectively, 100 and 400 solutions. Finally, we ran the algorithm for 1000 generations, which was verified to be sufficient for NSGA-II to converge to the Pareto front (i.e., we assessed that the results did not change significantly when running the method for more generations).

In what follows, we report the most relevant results obtained within the study. We also discuss some outcomes that illustrate a variety of possibilities offered by the framework of PFA, even if the suggested modifications of performances were judged unrealistic by the experts during the final feedback focus group. The analysis was performed for the individual DMs, for the average model as well as for the set of models $\mathcal{U}$ indicating the space in which the consensus between DM_1 and DM_2 can be searched.

5.2. Post Factum Analysis for the potentially optimal actions

In this subsection, we focus on the analysis of the potentially optimal actions, i.e. a_1 and a_3. They are possibly ranked first for at least one relevant value function in $\mathcal{U}$. Moreover, a_1 is ranked at the top for DM_2, whereas a_3 achieves the first position for DM_1 as well as for the average model U_{AVG}. In this perspective, it is relevant to consider how a_1 should improve to be ranked first for U_{DM_1} and U_{AVG} as well as how much better a_2 needs to perform to achieve the first rank for DM_2. Moreover, for both actions it is interesting to consider the improvements that they need to make for the necessary achievement of the first rank.

The results of such an investigation involving individual criteria are presented in Table 7. For example, to achieve the first rank for DM_1, a_1 needs to either decrease the length of construction works (g_3) by almost 9 months, or reduce the costs (g_5) by over 530 thousand euro, or increase the number of potential users (g_8) by over 65 thousands. Whatever the improvement on the remaining relevant criteria (g_1, g_4, g_6, and g_7), it does not allow a_1 achieving the first position in the ranking for DM_1. The results for the average model are analogous as for U_{DM_1}. However, the improvements on g_3, g_5, or g_8 to reach the first rank are significantly lower. For example, the duration of construction works (g_3) needs to be reduced by less than one month only (0.69) and the number of potential users (g_8) needs to be improved by just over 12000. These two scenarios were deemed realistic by the experts participating in the focus group as the initial forecast of a_1 on g_3 was a precautionary one, while the improvement on g_8 can be easily obtained with a proper communication campaign. Finally, no improvement on a single criterion can guarantee the first position to a_1 for all relevant models in $\mathcal{U}$.

For a_3 the necessary achievement of the first rank is possible, though solely with a significant improvement in terms of creation of new green areas (g_1). When it comes to the analysis for DM_2, a_3 would be ranked at the very top with a considerable improvement on g_1, g_4, or g_7. For example, the impact on the touristic sector (g_7) needs to be modified by one performance level from medium to high. Nevertheless, the experts agreed that the proposed improvements on g_1, g_4, or g_7 are not achievable by a_3.

Table 7: Improvements on the individual criteria that allow a_1 or a_3 achieving the first position in the ranking ('-' means that the target cannot be achieved).

Target	g_1	g_3	g_4	g_5	g_6	g_7	g_8
$T^{U_{DM_1}}(a_1 \rightarrow 1)$	-	8.98	-	530845.02	-	-	65513.39
$T^{U_{AVG}}(a_1 \rightarrow 1)$	-	0.69	-	155062.7	-	-	12836.14
$T^{\mathcal{U}}_N(a_1 \rightarrow 1)$	-	-	-	-	-	-	-
$T^{U_{DM_2}}(a_3 \rightarrow 1)$	30246.51	-	1	-	-	1	-
$T^{\mathcal{U}}_N(a_3 \rightarrow 1)$	120040.8	-	-	-	-	-	-

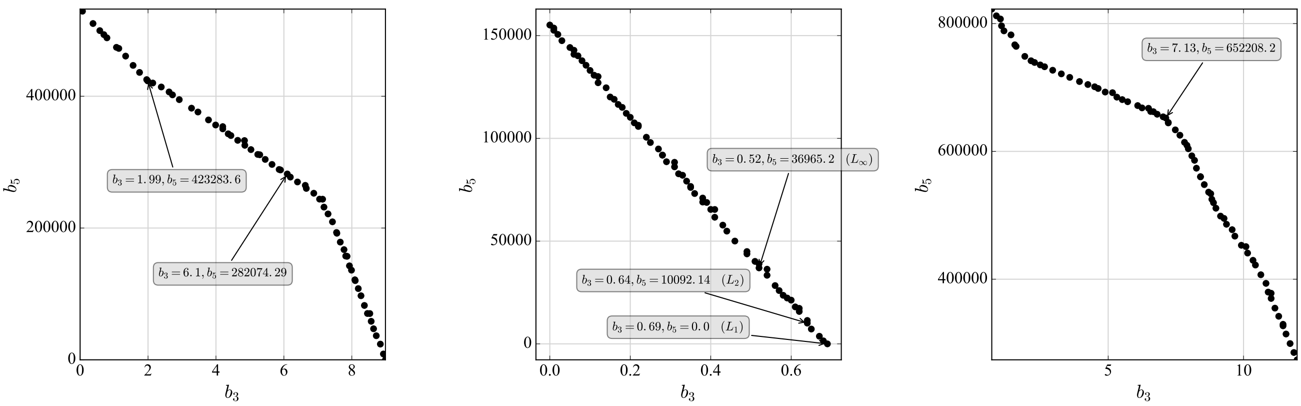
PFA for the potentially optimal actions has also focused on changing the performances on a few criteria simultaneously. The experts indicated the following triplet of criteria as particularly interesting for a_1: new green areas (g_1), duration of construction works (g_3), and costs (g_5).

Since a_1 achieves the maximal possible performance on g_1, we focus on the analysis of g_3 and g_5. The solutions which exhibit the trade-offs between the improvements required on these two criteria are presented in Figure 4. Let us emphasize that all figures present required/allowed modifications b_j of the current performances rather than the performances after these modification. By analyzing simultaneous improvements on g_3 and g_5, we adopt a similar perspective as in Data Envelopment Analysis (DEA) [9], i.e., we indicate which improvements need to be made on different criteria to achieve some better results. Once the solutions are computed, it is the task of the involved stakeholders to judge whether the suggested modifications are realistic. Obviously, one may claim that from a project management perspective compressing time requires increasing costs in a project. This would involve deteriorating performance of a_1 on one criterion and improving its performance on another criterion. Even though such a setting was not relevant for the stakeholders involved in our study in the context of g_3 and g_5, in the following sections we demonstrate that the analysis of deteriorating expenses and improving the performances on other criteria so that to reach some target can be handled by the proposed framework.

When comparing any pair of solutions in Figure 4 one can conclude that the same target can be achieved by a_1 with different performance changes, and a less modification on one criterion needs to be compensated with a greater modification on the other criterion. In particular, the two solutions marked in Figure 4a) exemplify that a_1 can achieve the same target either by reducing the duration of construction works (b_3) by just less than 2 months and the costs (b_5) by over 423 thousand euro, or by shortening the works by about 6 months and the costs by about 282 thousand euro (thus, respectively, more on g_3 and less on g_5).

Note that for DM_1 the extreme solutions correspond to single-criterion improvements already presented in Table 7, i.e., $[b_3, b_5]$ equal to $[8.98, 0]$ and $[0, 530845.02]$ (see Figure 4a). However, the remaining solutions indicate complex non-linear trade-offs between the required improvements. For example, when taking solution $[b_3, b_5] = [8.98, 0]$ as a reference, the expected decrease in the length of construction works can be reduced from $b_3 = 8.98$ to about 7 months though only when the overall expenses of implementing a_1 are decreased by about $b_5 = 250$ thousand euro (as compared to no expected reduction of costs $b_5 = 0$). Following the same reasoning, when referring to other extreme solution $[b_3, b_5] = [0, 530845.02]$, one can reduce the expenses by $b_5 = 400$ rather than 530 thousand euro when the construction works are shortened by 2.5 months (as compared to no required time compression $b_3 = 0$).

Conversely, for the average model the trade-offs between changes required on g_3 and g_5 are linear (see Figure 4b). Thus, starting from the solution $[b_3, b_5] = [0, 155062.7]$, the required reduction of costs (b_5) can be decreased by about 20 thousand euro provided that the length of construction works is reduced by about 3 days (i.e., 0.1 month).



(a) $T^{U_{DM_1}}(a_1 \to 1)$ (b) $T^{U_{AVG}}(a_1 \to 1)$ (c) $T^{\mathcal{U}}_N(a_1 \to 1)$

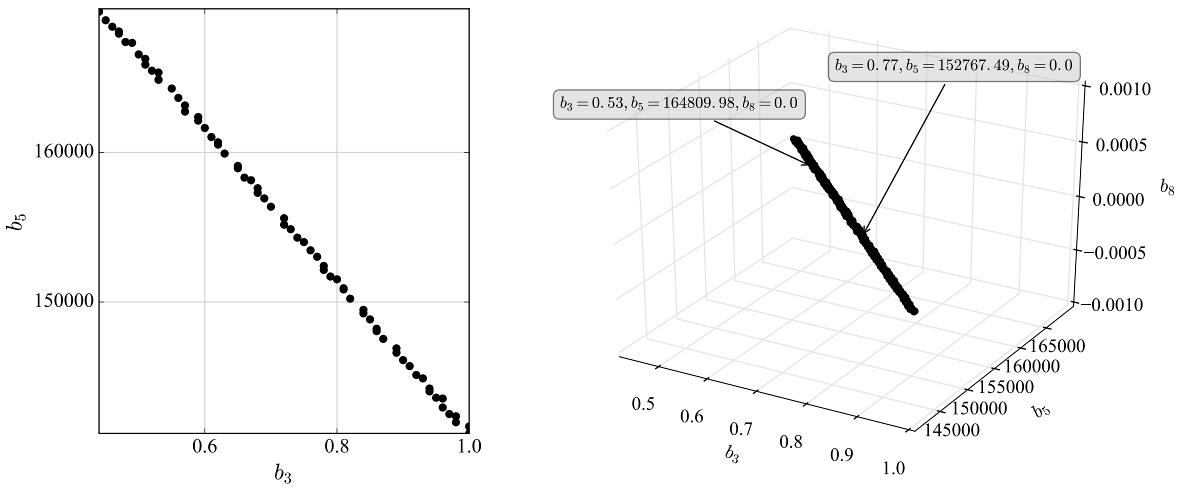
Figure 4: The solutions representing minimal improvements on g_3 and g_5 that allow a_1 achieving the first position in the ranking.

When it comes to the analysis of the set of all relevant models $\mathcal{U}$, a_1 can be necessarily ranked first if it simultaneously improves on g_3 and g_5 (see Figure 4c). The shape of the Pareto front representing the minimal improvements in this case is rather concave. The analysis of the possible trade-offs can be started with a solution located in the most concave part of the Pareto front which has been marked in Figure 4c. This solution corresponds to reducing the length of construction works by just over 7 months and the costs by about 650 thousand euro. Then, one can realize that with a slight further reduction on g_3 or g_5, the required improvement on the other criterion can be significantly limited.

Although the necessary changes on g_3 and g_5 for a_1 to become first for all relevant models and DM_1 were judged too significant by the experts, they found the analysis for the average model quite interesting. In particular, they acknowledged that a slight reduction of costs (g_5) is realistic when combined with the shortening of works by several days (g_3) (see Figure 4b). The reasoning behind this assumption might be that in territorial planning management the practice is usually to adopt a precautionary principle, i.e. multiplying estimated costs and duration of projects by some *safety coefficients*. In this perspective, let us note that alternative solutions contained in the Pareto front can be further exploited to indicate the scenarios which minimize some distance measure. Since performances on different criteria are expressed on various scales, some normalization of the performances should be conducted before computing the distances. In particular, we have normalized the performance modifications on each criterion with respect to the respective performance space. In this way, one can judge how large is the required change in view of the performances achieved by different actions. Then, the following vector of modifications optimizes the L_1-norm: $[b_3, b_5] = [0.69, 0.0]$, thus, minimizing the sum of normalized improvements that are required for a_3 on g_3 and g_5 to achieve the target.

Furthermore, $[b_3, b_5] = [0.69, 10092.14]$ and $[b_3, b_5] = [0.52, 36965.2]$ minimize the required improvements in terms of, respectively, L_2- and L_∞-norms (i.e., respectively, the Euclidean length of a vector representing all performance modifications and the maximal required performance change on some criterion).

As far as a_3 is concerned, the analysis focused on the improvements required on g_3, g_5, and g_8 so that it is ranked first. In particular, the results of PFA indicate that although a_3 cannot reach the first rank for DM_2 by improving its performance solely on either g_3 or g_5 (note that the minimal values of b_3 and b_5 in Figure 5a) are both greater than zero), it can indeed reach the first position with simultaneous improvements on both criteria. Nonetheless, these changes were judged unrealistic by the experts. Moreover, the improvement on g_8 was not contributing to the target achievement. The same has been observed when studying the necessary achievement of the first rank by a_3. To illustrate this, Figure 5b) exhibits the required performance changes involving all three criteria: g_3, g_5, and g_8. However, in all considered settings b_8 is equal to zero. Thus, to be ranked first for all feasible weight vectors a_3 needs to improve on g_3 and g_5 with the proviso that different trade-offs between the modifications on these criteria are allowed. To support their interpretation, we have marked some exemplary solutions in Figure 5b. Still, the experts judged that such significant reductions of costs and duration of construction works were not feasible for this option.



(a) b_3 and b_5 for $T^{U_{DM_2}}(a_3 \to 1)$

(b) b_3, b_5 and b_8 for $T^{\mathcal{U}}_N(a_3 \to 1)$

Figure 5: The solutions representing minimal improvements that allow a_3 achieving the first position.

For illustrative purpose, let us remind that a_1 is ranked second ($P_{*,\mathcal{U}}(a_2) = 2$) for its least advantageous value function in the set of possible consensus models $\mathcal{U}$. The position in the top two would be maintained for at least one value function with the following maximal deteriorations on the individual criteria: 165000m^2 on g_1, three performance levels on g_4, four classes and one class on, respectively, g_6 and g_7, or 75000 users on g_8. Furthermore, the last three scenarios would allow maintaining the position in top two also in the necessary sense. As for g_1 and g_4, the admissible deteriorations are smaller (119327.26m^2 and 1 performance level, respectively).

Bearing in mind that a_1 is necessarily preferred to all actions but a_3, the experts were also interested in the margin of safety that a_3 had preventing a_1 from being necessarily ranked first. This perspective differs from the standard one adopted in PFA as it investigates the modifications of performances of one action that allow another action achieving some target. In Figure 6, we present some exemplary results that were presented to the experts. These exhibit the minimal deteriorations of a_3 on g_5 and g_8 that allow a_1 to be necessarily ranked first. The experts concluded that such a margin of safety is large as both the increase in costs (g_5) or the decrease in the number of potential users (g_8) were significant.

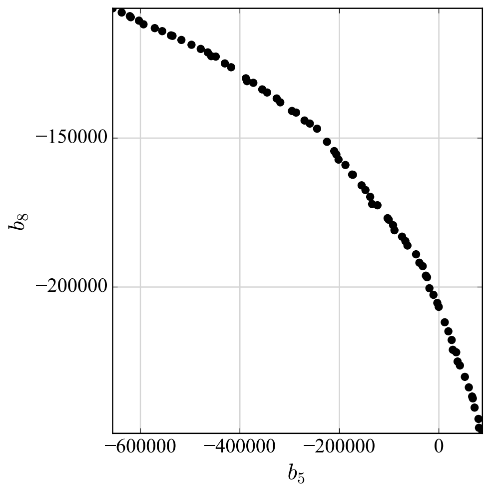


Figure 6: The solutions representing the minimal deteriorations of a_3 on g_5 and g_8 that allow a_1 being necessarily ranked first.

Both experts participating in the feedback focus group recognized the value of discovering what improvements are needed to make the actions under analysis reach the first place, not only for the other DM but also for the average model. There are a few reasons for this:

- knowledge about the needed improvements provides indications on how to improve/generate new actions/products/services;
- it allows a learning effect to take place as each DM becomes aware of small improvements that can change the final results (this was observed for the average model);
- it allows an enriched discussion about the feasibility of the performance improvements and can thus better support the achievement of a final consensus about which option to choose;
- discovering which improvements are needed can help to obtain better alternatives overall, which are going to benefit also the community.

The above considerations were reported by the experts involved in the focus group and recorded by the facilitator. The development of the discussion during the feedback focus group thus generated the cognitive-level impacts on the different participants (e.g., [29]).

5.3. Post Factum Analysis for the lower-ranked actions in terms of all relevant value functions

In this subsection, we focus on the rank-related targets concerning the lower-ranked actions (a_2 and a_4) and all relevant value functions. Let us remind the reader that a_4 has been ranked in positions between 2 and 4 (see Table 5). In this perspective, it is reasonable to consider the improvements that are required for a_4 to achieve different positions at the podium. The results of the analysis involving individual criteria are presented in Table 8. For the possible scenario, no improvement is required for the second and third ranks as these are achievable already with the current performances. Interestingly, a_4 can be possibly ranked at the top when improving its performance on one out of four criteria. The possible scenarios involve the creation of additional green areas (g_1), or the reduction of the length of the construction works (g_3), or the reduction of the costs (g_5), or the increase of the number of potential users (g_8). However, these improvements were all judged as too significant by the experts participating in the final focus group as these were at least doubling the original performances. In fact, DM_2 explained that it was not surprising that it is difficult for a_4 to be ranked first with a single criterion improvement, because it is indeed different from the other options. It represents a punctual intervention whereas other actions correspond to linear interventions across the whole railway line.

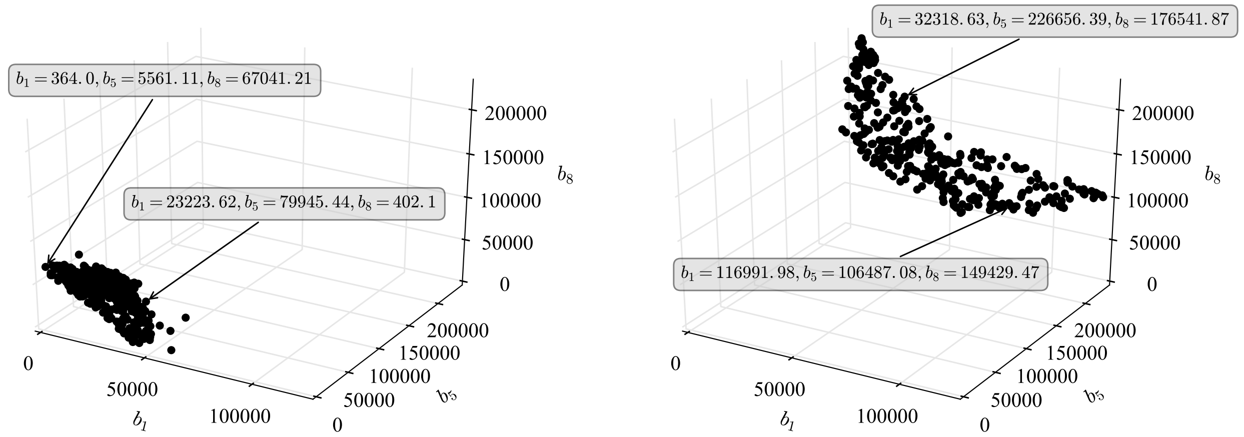
The same reaction of the experts was raised by the analysis of the improvements that would guarantee that a_4 is necessarily ranked at the podium. Although marginal improvements on four criteria would be required for the necessary achievement of the third rank by a_4, no single criterion improvement could ensure that it is necessarily judged as one of the top two options.

Table 8: Improvements on the individual criteria that allow a_4 achieving a position at the podium in the possible or necessary sense ('-' means that the target cannot be achieved).

Target	g_1	g_3	g_4	g_5	g_6	g_7	g_8
$T_P^{\mathcal{U}}(a_4 \to 1)$	46884.1	2.58	-	115564.93	-	-	67390.56
$T_N^{\mathcal{U}}(a_4 \to 1)$	-	-	-	-	-	-	-
$T_P^{\mathcal{U}}(a_4 \to 2)$	0	0	0	0	0	0	0
$T_N^{\mathcal{U}}(a_4 \to 2)$	-	-	-	-	-	-	-
$T_P^{\mathcal{U}}(a_4 \to 3)$	0	0	0	0	0	0	0
$T_N^{\mathcal{U}}(a_4 \to 3)$	5652.25	0.26	-	22959.73	-	-	4868.57

To investigate simultaneous improvements on a few criteria, the experts indicated the following triplet as the most interesting one: new green areas (g_1), costs (g_5), and number of potential users (g_8). In Figure 7, we present the alternative solutions that allow a_4 achieving the first rank in the possible and necessary sense. These illustrate how significant are the differences in the required improvements when considering the two scenarios. For example, when accounting for the achievement of the first rank by a_4 for at least one value function, with a marginal reduction of costs (g_5) there are plenty of scenarios involving a variety of trade-offs between b_1 and b_8. On the contrary, when opting for a negligible reduction of costs, a_4 could be necessarily ranked first only with the major improvements on both g_1 and g_8.

The experts noted that the improvements needed for the necessary achievement of the first rank by a_4 are unlikely to happen. At the same time, they both confirmed that the changes required to implement the possible scenario were manageable. However, they expressed doubts on whether the reduction of costs is possible with further improvements of the performances on g_1 or g_8. For this reason, they expressed the interest in investigating the scenarios in which additional new green areas are created or the number of potential users is increased at the additional cost. To investigate this, we relaxed the constraint of considering only $b_j \geq 0$, $j = 1, 5, 8$, thus, admitting potential deterioration of the current performances.

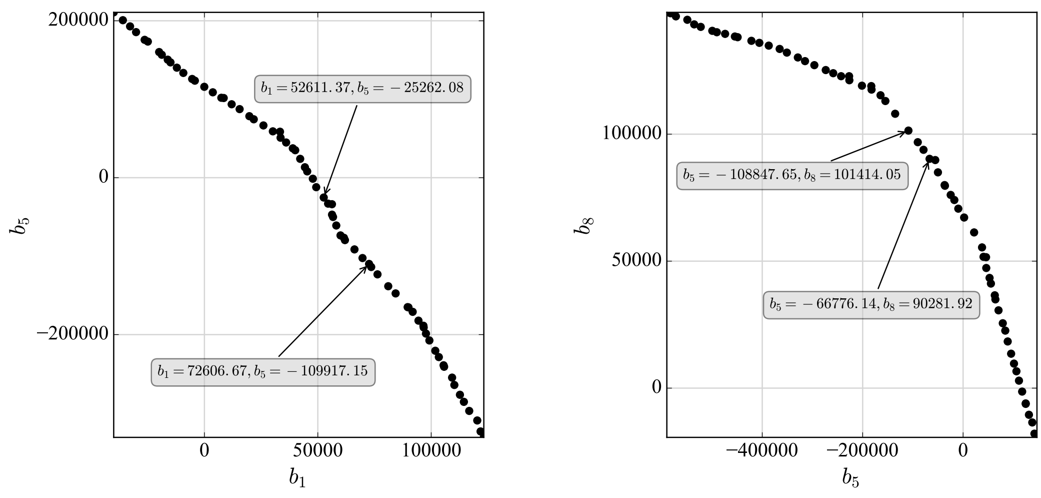


(a) Possible

(b) Necessary

Figure 7: The solutions representing minimal improvements on g_1, g_5 and g_8 that allow a_4 achieving the first rank.

In Figure 8, we present the modifications of performances allowing a_4 to necessarily achieve the first rank. Note that $b_5 < 0$ corresponds to the increase in costs. In fact, the experts participating in the feedback focus group acknowledged that it would be possible to create more green areas or attract more users with a communication campaign when additional funds are available. Some interesting feasible solutions have been marked in Figure 8. For example, a_4 would be necessarily ranked first if it involved the creation of about $52611m^2$ of additional new green areas for at most 25262 euro that would increase the overall costs of implementing this action (see $[b_1, b_5] = [52611.37, -25262.08]$). These scenarios were deemed realistic, and overall their feasibility increased the attractiveness of a_4 for the experts.



(a) g_1 and g_5

(b) g_5 and g_8

Figure 8: The solutions representing the modifications of performances allowing a_4 to necessarily achieve the first rank.

To further confirm the superiority of a_4 over a_2 and a_5, we investigated the conditions under which a_4 was necessarily ranked at the podium (thus, possibly worse only than a_1 and a_3). For this purpose, the experts indicated that the number of attractors could change if the extension of green areas was involved. Thus, we investigated the changes on g_9, highlighting that they are linked to the modifications on g_1. In Figure 9, we present the minimal improvements required on g_1 and g_9 so that a_4 is necessarily ranked third. The required changes are marginal and were judged as realistic to be implemented. Based on all favourable results of PFA for a_4, DM_2 confirmed that should a sufficient budget become available, a_4 could be developed in conjunction with a_1.

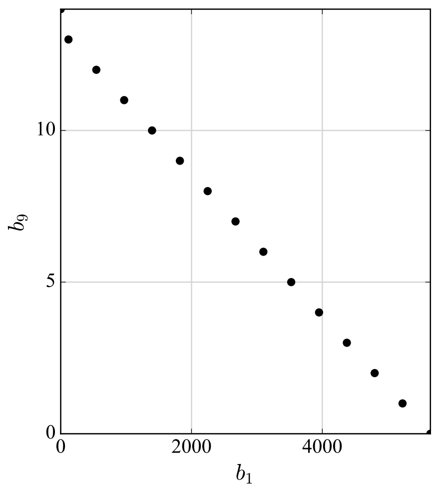


Figure 9: The solutions representing minimal improvements on g_1 and g_9 that allow a_4 to necessarily achieve the third position.

To further illustrate different scenarios that can be derived from Post Factum Analysis, let us focus on a_2. Since with the current performances it is necessarily ranked fifth, we investigated how it should improve to achieve better positions. The results for the single criteria improvements are presented in Table 9. These indicate that:

- Action a_2 cannot achieve the first rank in the possible sense nor any position at the podium in the necessary sense.
- The less demanding the target, the greater the number of criteria that can be individually improved so that a_2 can achieve better positions. For example, a_2 can be possibly ranked second when improving significantly either only on g_1 or g_8, whereas the scenarios for the possible achievement of the third rank already involve individual improvements on five criteria: g_1, g_4, g_5, g_7, and g_8. Following the same reasoning, a_2 can be possibly ranked fourth with an adequately great modification on either g_6 or g_7 (see Table 9), while no improvement on these criteria guarantees achieving the target in the necessary sense.
- The more demanding the target, the greater the improvement required to achieve it. For example, a_2 needs to improve its performance on g_8 by 13106.2, 59639.18, or 146175.71 users to possibly reach, respectively, the fourth, third, and second rank; analogously, the costs (g_5) related to a_2 need to be reduced by 52742.52 or 146402.21 euro so that it is ranked fourth in the possible or necessary sense, respectively.

The experts agreed that the only realistic improvement achievable for a_2 is the one on g_4 (impact on the landscape). Thus, they concluded that it would be feasible for a_2 to be possibly ranked third, but not higher. The infeasibility of achieving the first two positions by improving some individual performance by a_2, strengthened the perception of a_1 and a_3 as the best options by the experts.

Table 9: Improvements on the individual criteria that allow a_2 achieving different ranks in the possible or necessary sense ('-' means that the target cannot be achieved).

Target	g_1	g_3	g_4	g_5	g_6	g_7	g_8
$T_P^{\mathcal{U}}(a_2 \to 1)$	-	-	-	-	-	-	-
$T_N^{\mathcal{U}}(a_2 \to 1)$	-	-	-	-	-	-	-
$T_P^{\mathcal{U}}(a_2 \to 2)$	159792.37	-	-	-	-	-	146175.71
$T_N^{\mathcal{U}}(a_2 \to 2)$	-	-	-	-	-	-	-
$T_P^{\mathcal{U}}(a_2 \to 3)$	47991.75	-	1	142700.61	-	1	59639.18
$T_N^{\mathcal{U}}(a_2 \to 3)$	-	-	-	-	-	-	-
$T_P^{\mathcal{U}}(a_2 \to 4)$	8292.65	-	1	52742.52	2	1	13106.2
$T_N^{\mathcal{U}}(a_2 \to 4)$	67411.34	-	1	146406.21	-	-	99344.55

The latter conclusion was confirmed with the analysis of improvement scenarios involving simultaneously g_1, g_5, and g_8. The solutions representing minimal required improvements for the possible achievement of different ranks are presented in Figure 10. These clearly show how significantly a_2 needs to improve in order to obtain the respective better positions.

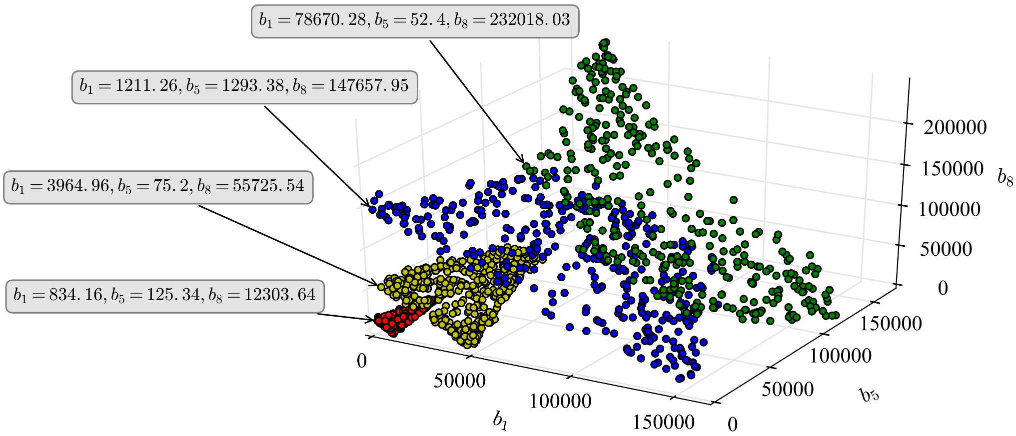


Figure 10: The solutions representing minimal improvements on g_1, g_5 and g_8 that allow a_2 possibly achieving different ranks: red – 4, yellow – 3, blue – 2, and green – 1.

5.4. Post Factum Analysis for the targets involving pairs of actions

In this subsection, we present exemplary results concerning preference-related targets. These concern the truth of preference relation for some pairs of actions. Let us emphasize that the outcomes presented in Section 5.2 concerning achievement of the first rank by a_1 and a_3 are the same as for the direct comparison of these two actions. It means that the conditions under which a_1 or a_3 are ranked at the top are equivalent to the ones that guarantee they are preferred to each other. Thus, in what follows we compare these actions against a_4.

Let us first observe that a_3 is possibly though not necessarily preferred to a_4. Thus, it is relevant to verify how a_3 needs to improve so that the preference is valid for all models under consideration. The results presented in Table 10 indicate that this can happen with a slight improvement on one of the six criteria (g_1, g_3, g_4, g_5, g_6, or g_7). For example, in terms of the ordinal criteria, a_3 needs to improve the impact either on the landscape (g_4) from irrelevant to positive or on the touristic sector (g_7) from medium to high. The experts judged that only the improvements on g_5 and g_6 were realistic in the short term. Obviously, when considering a few criteria simultaneously, one can reduce the required improvement on some criteria at the cost of improving on other criteria. This is illustrated with Figure 11 which exhibits the trade-offs between required changes on g_1, g_3, and g_5.

Table 10: Improvements on the individual criteria that allow a_3 being necessarily preferred to a_4 ('-' means that the target cannot be achieved).

Target	g_1	g_3	g_4	g_5	g_6	g_7	g_8
$T_N^{\mathcal{U}}(a_3 \succsim a_4)$	3625.28	0.75	1	23057.31	1	1	-

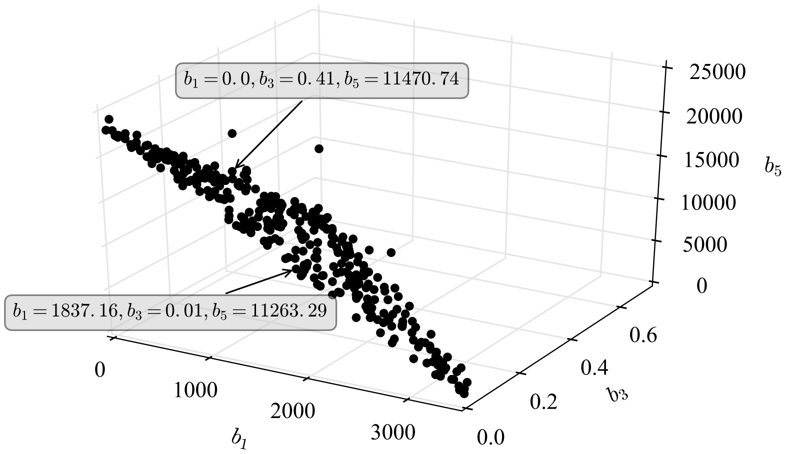


Figure 11: The solutions representing minimal improvements on g_1, g_3 and g_5 that allow a_3 being necessarily preferred to a_4.

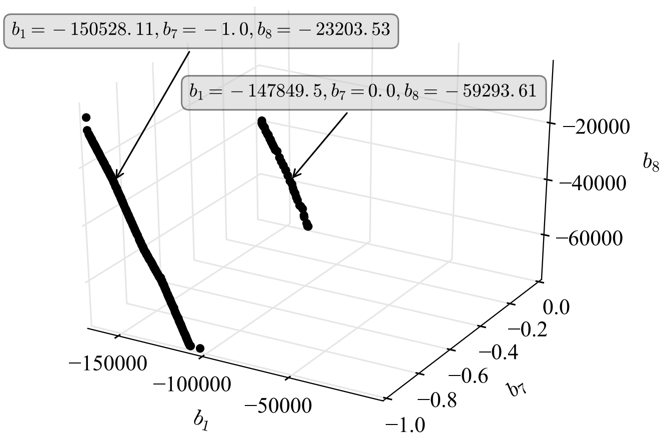
When it comes to a_1, this action is possibly and necessarily preferred to a_4. Thus, in this case one can focus on the deteriorations that a_1 can afford so that it is still preferred to a_4. When it comes to deteriorations on the individual criteria, the possible relation would be maintained even with a considerable deterioration on g_1, g_4, g_6, g_7, or g_8 (see Table 11). Note, however, that no reduction is allowed on g_3 and g_5. For the necessary relation, the deterioration on g_4 and g_7 also cannot be afforded, and the margin of safety on the remaining criteria is much lower. For example, $a_1 \succsim^P a_4$ with the reduction of new jobs from 4 to 0, but $a_1 \succsim^N a_4$ would hold only if the number of new jobs is at least 2. The experts found it extremely relevant to discover what are the critical criteria on which a_1 cannot afford any decrease of the performance to maintain its advantage over a_4.

Table 11: Deteriorations on the individual criteria that allow a_1 to be possibly or necessarily preferred to a_4.

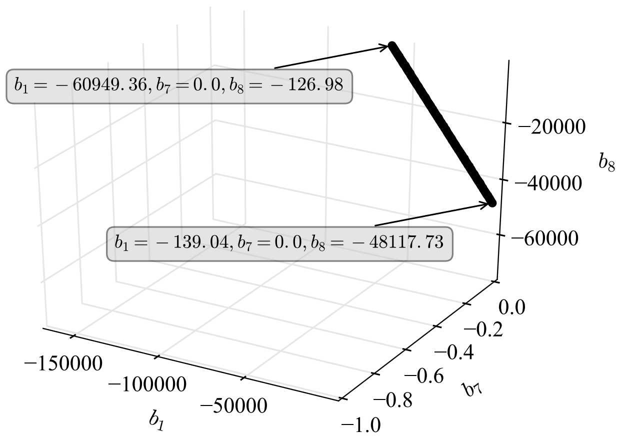
	g_1	g_3	g_4	g_5	g_6	g_7	g_8
$T_P^{\mathcal{U}}(a_1 \succsim a_4)$	-165000.0	0	-3	0	-4	-1	-75000.0
$T_N^{\mathcal{U}}(a_1 \succsim a_4)$	-61171.89	0	0	0	-2	0	-48310.1

To illustrate that the investigation of the simultaneous changes on a few criteria could also involve the ordinal ones, let us focus on g_1, g_7, and g_8 (with the performances on g_7 being expressed on the ordinal scale). The solutions representing the maximal deteriorations on g_1, g_7 and g_8 that allow a_1 being preferred to a_4 in the possible and necessary sense are presented in Figure 12.

When not admitting deterioration on g_7 ($b_7 = 0$), the trade-off between deteriorations on g_1 and g_8 allowing the maintenance of the possible relation exhibits a linear trend. The same kind of interrelations between b_1 and b_8 can be observed for $b_7 = -1.0$. However, the allowed reductions on g_1 and g_8 are significantly less when compared with the case of $b_7 = 0$ (see Figure 12a). Finally, to keep the necessary relation true one cannot tolerate any deterioration on g_7 (see Figure 12b). Also, the reductions admitted on g_1 and g_8 are significantly lower than in the case of the target involving the possible preference.



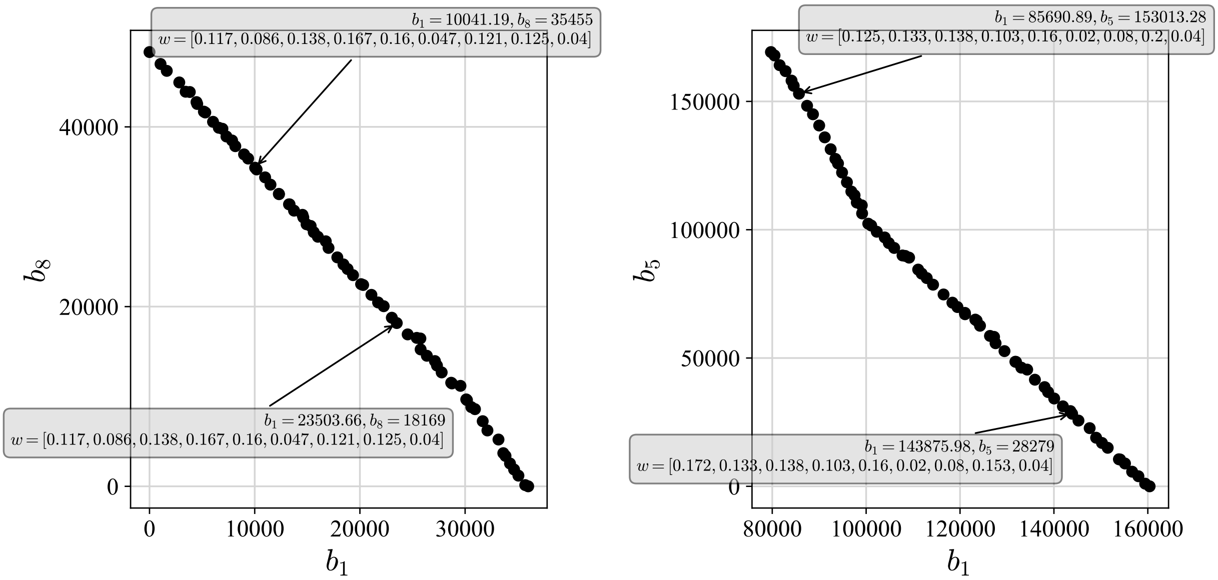
(a) Possible preference



(b) Necessary preference

Figure 12: The solutions representing maximal deteriorations on g_1, g_7 and g_8 that allow a_1 being preferred to a_4.

For illustrative purposes, in Figure 13 we present the minimal performance improvements that allow either a_4 or a_2 being possibly preferred to a_1. For some selected solutions, we explicitly provide the weight vectors for which the target would be achieved with the corresponding performance modifications. In Figure 13a), these weights are the same for a pair of marked solutions. However, in Figure 13b) the provided weights do differ with respect to w_1 and w_8. This proves that each solution is associated with some compatible weight vector that is the most advantageous in terms of achieving the target in the possible sense.



(a) $T_P^{\mathcal{U}}(a_4 \succsim a_1)$ (b) $T_P^{\mathcal{U}}(a_2 \succsim a_1)$

Figure 13: The solutions representing minimal performance improvements on a pair of criteria that allow one action being possibly preferred to another action.

6. Conclusions

The requalification of an abandoned railway line is a complex decision making problem involving multiple and conflicting perspectives (see, e.g., [18, 63]). In this paper, we took into account the preferences of representatives of a public entity and a private organization to identify the most advantageous options for the requalification of an abandoned railway line in the North of Italy. Such a recommendation was used as an input within a framework of Post Factum Analysis that considered the impact of performance changes on the obtained results.

From the methodological point of view, in this paper we improved Post Factum Analysis by formulating a Multiple Objective Optimization problem in which the performance changes on the individual criteria are simultaneously minimized to allow the action achieving certain targets. Moreover, we made the framework more usable by considering different performance scales and exhibiting trade-offs between the required/allowed changes on different criteria.

The main focus of the performed analysis was on the minimal improvement of actions' performances that would ensure feasibility of some currently impossible outcomes and on the maximal deterioration that would allow actions maintaining some already achieved targets. We also demonstrated that Post Factum Analysis can be used in the following realistic contexts:

- to allow performance deterioration on a criterion (e.g., the increase in costs) that would justify feasibility of required improvements on some other criteria;
- to investigate the modification of performances of one option that would allow another option achieving certain targets.

The results proposed in Section 5 were shown and discussed with two experts representing DM_1 and DM_2 in a final feedback focus group facilitated by the second author of this paper. The following paragraphs summarise the final comments provided by the participants on the proposed tools and analyses from the point of view of their operability, consensus building capacity and transferability. In particular, both DMs appreciated the purpose of the proposed analyses/simulations and recognised the value and the merit of the tools.

Post Factum Analysis offered the arguments in favour or against the respective options under analysis, thus, enabling a better informed discussion about the results among the participants in the process. Indeed, the experts identified a_1 and a_3 as the most advantageous actions, a_2 and a_5 as under-performing options, and a_4 as a promising alternative. Both DMs stated that the obtained overall results are coherent with their expectations. They both better appreciated why $a_1 + a_4$ could represent a very interesting solution in this decision making context. Since there are 12 abandoned railway lines in the considered region, they started discussing the idea that for the specific railway under analysis in this paper it might be an indeed interesting solution to have $a_1 + a_4$ given the mixed urban and rural territorial context crossed by the railway, thus leading to a significant touristic potential, whereas for some other abandoned railway lines a_3 could be the best option and for some others even a_2 could be an interesting solution. These findings illustrate how Post Factum analysis can support the development of new/better alternatives. They also provide the motivation for further developments of this research, by extending the analyses proposed in [18] and in this paper to the remaining abandoned railway lines in the Piedmont Region.

The experts found Post Factum Analysis understandable and transferable to future applications. On one hand, DM_2 noticed that the proposed tool supported negotiation and legitimation of the final results which had to be communicated to the public. Thus, she was interested in employing it in collaborative projects to help all stakeholders understand the dynamics between different actions. On the other hand, DM_1 stated that he would be interested in using the framework for decision making problems that are characterized by high imprecision with respect to the actions' performances (to provide space for realistic simulations about improvements and deteriorations). Let us emphasize that in the latter case, such an imprecision represented with, e.g., the probability distribution on the space of possible performances [45] or n-point intervals [12], can be also accounted for within the framework of robustness analysis. Then, the stability of recommendation may be again quantified with the stochastic acceptability indices [45] or the necessary preference relation [12] that would reflect not only the plurality of compatible preference model instances, but also different feasible performances of actions.

With reference to the limitations, the experts concluded that in some applications keeping the bounds of possible changes within the original performances ranges may be too restrictive, making it difficult to think about new actions which might not have been foreseen before. To overcome this, the preference model (e.g., marginal value functions) need to be elicited for the whole range of performances that is found relevant for Post Factum Analysis. Moreover, from the research design point of view, we acknowledge that the discussion of a single case study limits the possibility of generalising the obtained results [21, 67].

We envisage the following future developments of the proposed framework motivated by the indications of the experts involved in the study. First, we will propose a stepwise benchmarking tool [51] for Post Factum Analysis that would indicate an optimal development path toward the target. Secondly, we will develop some visualization tools to show trade-offs between the required/allowed performances changes on different criteria on the interactive bi- or tri-dimensional plots. Such tools could be used in real time during the collaborative focus groups, thus, increasing the applicability of Post Factum Analysis in practice (see [19] for the available meta-choices involved with the design of visualization tools within collaborative settings). Thirdly, we will make PFA usable with outranking- (see, e.g., [7, 20]) and rule-based (see, e.g., [8, 25, 28]) preference models. Finally, the methodological framework will be extended to the context of Portfolio Decision Analysis [57] where a subset of the most preferred actions needs to be selected. This requires consideration of the performance modifications of more than one action at the same time.

Acknowledgements

We thank the participants in the focus group led by the second author of this paper where the results of the Post Factum Analysis were analysed and discussed for the crucial insights and feedback provided on the analysis. We also thank three anonymous referees for their remarks which helped us to significantly improve the paper. Krzysztof Ciomek and Miłosz Kadziński acknowledge financial support from the Polish Ministry of Science and Higher Education under the Iuventus Plus program in 2016-2019 Grant number IP2015 029674 - 0296/IP2/2016/74.

References

[1] Angilella, S., Corrente, S., Greco, S., 2015. Stochastic multiobjective acceptability analysis for the Choquet integral preference model and the scale construction problem. European Journal of Operational Research 240 (1), 172 – 182.

[2] Beinat, E. (Ed.), 1997. Value Functions for Environmental Management. Kluwer Academic Publishers, Dordrecht.

[3] Beynon, M. J., Wells, P., 2008. The lean improvement of the chemical emissions of motor vehicles based on preference ranking: A PROMETHEE uncertainty analysis. Omega 36 (3), 384 – 394.

[4] Bisdorff, R., Dias, L., Meyer, P., Mousseau, V., Pirlot, M. (Eds.), 2015. Evaluation and Decision Models with Multiple Criteria. International Handbooks on Information Systems. Springer, Berlin.

[5] Bous, G., Fortemps, P., Glineur, F., Pirlot, M., 2010. 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.

[6] Branke, J., Deb, K., Miettinen, K., Słowiński, R. (Eds.), 2008. Multiobjective Optimization: Interactive and Evolutionary Approaches. Springer, Berlin.

[7] Certa, A., Enea, M., Lupo, T., 2013. ELECTRE III to dynamically support the decision maker about the periodic replacements configurations for a multi-component system. Decision Support Systems 55 (1), 126 – 134.

[8] Chakhar, S., Saad, I., 2012. Dominance-based rough set approach for groups in multicriteria classification problems. Decision Support Systems 54 (1), 372 – 380.

[9] Charnes, A., Cooper, W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2 (6), 429 – 444.

[10] Ciomek, K., Kadziński, M., Tervonen, T., 2017. Heuristics for prioritizing pair-wise elicitation questions with additive multi-attribute value models. Omega 71, 27 – 45.

[11] Corrente, S., Greco, S., Słowiński, R., 2012. Multiple criteria hierarchy process in robust ordinal regression. Decision Support Systems 53 (2), 660–674.

[12] Corrente, S., Greco, S., Słowiński, R., 2017. Handling imprecise evaluations in multiple criteria decision aiding and robust ordinal regression by n-point intervals. Fuzzy Optimization and Decision Making 16 (2), 127–157.

[13] Damart, S., Dias, L., Mousseau, V., 2007. Supporting groups in sorting decisions: Methodology and use of a multi-criteria aggregation/disaggregation DSS. Decision Support Systems 43 (4), 1464 – 1475.

[14] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., Apr. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6 (2), 182–197.

[15] Doumpos, M., Zopounidis, C., Galariotis, E., 2014. Inferring robust decision models in multicriteria classification problems: An experimental analysis. European Journal of Operational Research 236 (2), 601 – 611.

[16] Durbach, I., 2009. The use of the SMAA acceptability index in descriptive decision analysis. European Journal of Operational Research 196 (3), 1229–1237.

[17] Ferretti, V., 2016. From stakeholders analysis to cognitive mapping and multi-attribute value theory: An integrated approach for policy support. European Journal of Operational Research 253 (2), 524 – 541.

[18] Ferretti, V., Degioanni, A., 2017. How to support the design and evaluation of redevelopment projects for disused railways? A methodological proposal and key lessons learned. Transportation Research Part D: Transport and Environment 52 (A), 29–48.

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

[20] Giannoulis, C., Ishizaka, A., 2010. A web-based decision support system with ELECTRE III for a personalised ranking of British universities. Decision Support Systems 48 (3), 488 – 497.

[21] Gibbert, M., Ruigrok, W., Wicki, B., 2008. What passes as a rigorous case study? Strategic Management Journal 29 (13), 1465–1474.

[22] Greco, S., Ehrgott, M., Figueira, J. (Eds.), 2016. Multiple Criteria Decision Analysis: State of the Art Surveys. International Series in Operations Research & Management Science, Vol. 233. Springer, Berlin.

[23] Greco, S., Kadziński, M., Mousseau, V., Słowiński, R., 2012. Robust ordinal regression for multiple criteria group decision problems: UTA$^{\mathrm{GMS}}$-GROUP and UTADIS$^{\mathrm{GMS}}$-GROUP. Decision Support Systems 52 (3), 549–561.

[24] Greco, S., Kadziński, M., Słowiński, R., 2011. Selection of a representative value function in robust multiple criteria sorting. Computers & Operations Research 38 (11), 1620 – 1637.

[25] Greco, S., Matarazzo, B., Słowiński, R., 2001. Rough sets theory for multicriteria decision analysis. European Journal of Operational Research 129 (1), 1 – 47.

[26] Greco, S., Matarazzo, B., Slowinski, R., 2004. Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. European Journal of Operational Research 158 (2), 271–292.

[27] Greco, S., Mousseau, V., Słowiński, R., 2008. Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. European Journal of Operational Research 191 (2), 416–436.

[28] Greco, S., Słowiński, R., Zielniewicz, P., 2013. Putting dominance-based rough set approach and robust ordinal regression together. Decision Support Systems 54 (2), 891–903.

[29] Henao, F., Franco, L., 2016. Unpacking multimethodology: Impacts of a community development intervention. European Journal of Operational Research 253 (3), 681 – 696.

[30] Hwang, C.-L., Lai, Y.-J., Liu, T.-Y., 1993. A new approach for multiple objective decision making. Computers & Operations Research 20 (8), 889 – 899.

[31] Hyde, K., Maier, H., 2006. Distance-based and stochastic uncertainty analysis for multi-criteria decision analysis in Excel using Visual Basic for Applications. Environmental Modelling & Software 21 (12), 1695 – 1710.

[32] Hyde, K., Maier, H., Colby, C., 2005. A distance-based uncertainty analysis approach to multi-criteria decision analysis for water resource decision making. Journal of Environmental Management 77 (4), 278 – 290.

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

[34] Jacquet-Lagrèze, E., Siskos, Y., 2001. Preference disaggregation: 20 years of MCDA experience. European Journal of Operational Research 130 (2), 233–245.

[35] Kadziński, M., Cinelli, M., Ciomek, K., Coles, S., Nadagouda, M., Varma, R., Kirwan, K., 2016. Co-constructive development of a green chemistry-based model for the assessment of nanoparticles synthesis. European Journal of Operational Research (in press).

[36] Kadziński, M., Ciomek, K., Rychły, P., Słowiński, R., 2016. Post factum analysis for robust multiple criteria ranking and sorting. Journal of Global Optimization 65 (3), 531–562.

[37] Kadziński, M., Ghaderi, M., Wasikowski, J., Agell, N., 2017. Expressiveness and robustness measures for the evaluation of an additive value function in multiple criteria preference disaggregation methods: An experimental analysis. Computers & Operations Research 87, 146 – 164.

[38] Kadziński, M., Greco, S., Słowiński, R., 2012. Extreme ranking analysis in robust ordinal regression. Omega 40 (4), 488–501.

[39] Kadziński, M., Tervonen, T., 2013. Robust multi-criteria ranking with additive value models and holistic pair-wise preference state-

ments. European Journal of Operational Research 228 (1), 169–180.

[40] Kadziński, M., Tervonen, T., 2013. Stochastic ordinal regression for multiple criteria sorting problems. Decision Support Systems 55 (1), 55 – 66.

[41] Kadziński, M., Tomczyk, M., 2017. Interactive evolutionary multiple objective optimization for group decision incorporating value-based preference disaggregation methods. Group Decision and Negotiation 26, 693–728.

[42] Keeney, R., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Trade-offs. Cambridge University Press, Cambridge.

[43] Knuth, D. (Ed.), 1998. The Art of Computer Programming: Volume 3: Sorting and Searching. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA.

[44] Lahdelma, R., Salminen, P., 2001. SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. Operations Research 49 (3), 444–454.

[45] Lahdelma, R., Salminen, P., 2010. Stochastic Multicriteria Acceptability Analysis (SMAA). Springer, Boston, pp. 285–315.

[46] Leskinen, P., Viitanen, J., Kangas, A., Kangas, J., 2006. Alternatives to incorporate uncertainty and risk attitude in multicriteria evaluation of forest plans. Forest Science 52 (3), 304–314.

[47] Lewandowski, A., Wierzbicki, A. (Eds.), 1989. Aspiration Based Decision Support Systems: Theory, Software and Applications. Springer, Berlin.

[48] McKay, J., Marshall, P., 2001. The dual imperatives of action research. Information Technology & People 14 (1), 46–59.

[49] Mersmann, O., 2014. mco: Multiple Criteria Optimization Algorithms and Related Functions. R package version 1.0-15.1.

[50] Montibeller, G., 2007. Action-Researching MCDA Interventions. In: Proceeding of the 49th Annual Conference of the Operational Research Society (OR 49). pp. 106–123.

[51] Petrovic, M., Bojkovic, N., Anic, I., Stamenkovic, M., Tarle, S. P., 2014. An electre-based decision aid tool for stepwise benchmarking: An application over EU Digital Agenda targets. Decision Support Systems 59, 230 – 241.

[52] Ravalico, J., Dandy, G., Maier., H., 2010. Management Option Rank Equivalence (MORE) – A new method of sensitivity analysis for decision-making. Environmental Modelling & Software 25, 171–181.

[53] Rizzon, B., Galichet, S., Clivill, V., 2014. Robustness threshold methodology for multicriteria based ranking using imprecise data. In: 2014 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making (MCDM). pp. 1–8.

[54] Rocco, C., Tarantola, S., 2014. Evaluating ranking robustness in multi-indicator uncertain matrices: an application based on simulation and global sensitivity analysis. In: Bruggemann, R., Carlsen, L., Jochen, W. (Eds.), Multi-indicator Systems and Modelling in Partial Order. Springer, Berlin, pp. 275–292.

[55] Roy, B., 2010. Two conceptions of decision aiding. International Journal of Multicriteria Decision Making 1(1), 74–79.

[56] Salo, A., Hamalainen, R., 2001. Preference ratios in multiattribute evaluation (prime)-elicitation and decision procedures under incomplete information. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 31 (6), 533–545.

[57] Salo, A., Keisler, J., Morton, A. (Eds.), 2011. Portfolio Decision Analysis: Improved Methods for Resource Allocation. Springer, New York.

[58] Siggelkow, N., 2007. Persuasion with case studies. The Academy of Management Journal 50 (1), 20–24.

[59] Stewart, T., 2005. Dealing with Uncertainties in MCDA. In: Figueira, J., Greco, S., Ehrgott, M. (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys. Springer, Boston, pp. 445–470.

[60] Tervonen, T., Figueira, J., 2008. A survey on stochastic multicriteria acceptability analysis methods. Journal of Multi-Criteria Decision Analysis 15 (1–2), 1–14.

[61] Tervonen, T., van Valkenhoef, G., Basturk, N., Postmus, D., 2013. Hit-And-Run enables efficient weight generation for simulation-based multiple criteria decision analysis. European Journal of Operational Research 224 (3), 552 – 559.

[62] Triantaphyllou, E., Sánchez, A., 1997. A sensitivity analysis approach for some deterministic multi-criteria decision-making methods. Decision Sciences 28 (1), 151–194.

[63] Tsamboulas, D., Mikroudis, G., 2006. TRANS-POL: A mediator between transportation models and decision makers' policies. Decision Support Systems 42 (2), 879 – 897.

[64] Vincke, P., 1999. Robust solutions and methods in decision-aid. Journal of Multi-Criteria Decision Analysis 8 (3), 181–187.

[65] Wallace, S., 2000. Decision making under uncertainty: Is sensitivity analysis of any use? Operations Research 48 (1), 20–25.

[66] Wolters, W., Mareschal., B., 1995. Novel types of sensitivity analysis for additive MCDM methods. European Journal of Operational Research 81, 281–290.

[67] Yin, R., 2008. Case Study Research: Design and Methods, fourth edition. Edition. Sage Publications, London.

[68] Zopounidis, C., Doumpos, M., 2002. Multicriteria classification and sorting methods: A literature review. European Journal of Operational Research 138 (2), 229 – 246.

[69] Zopounidis, C., Pardalos, P. (Eds.), 2010. Handbook of Multicriteria Analysis. Springer, Berlin.