Accepted Manuscript

A multi-criteria approach to sort and rank policies based on Delphi qualitative assessments and ELECTRE TRI: the case of smart grids in Brazil

Luis C. Dias , Carlos Henggeler Antunes , Guilherme Dantas , Nivalde de Castro , Lucca Zamboni

Please cite this article as: Luis C. Dias , Carlos Henggeler Antunes , Guilherme Dantas , Nivalde de Castro , Lucca Zamboni , A multi-criteria approach to sort and rank policies based on Delphi qualitative assessments and ELECTRE TRI: the case of smart grids in Brazil, *Omega* (2017), doi: 10.1016/j.omega.2017.04.004

Highlights

- ELECTRE TRI-based sorting methodology for Delphi elicitation processes

- Qualitative assessments from a group translated into weight constraints

- Sorting-ranking methodology based on references (category profiles)

- Conjoint use of multi-criteria stochastic and robustness analyses

- Policies for smart grids in Brazil assessed under multiple perspectives

A multi-criteria approach to sort and rank policies based on Delphi qualitative assessments and ELECTRE TRI: the case of smart grids in Brazil

Luis C. Dias [a,b], Carlos Henggeler Antunes [a,c], Guilherme Dantas [d], Nivalde de Castro [d], Lucca Zamboni [e]

[a] *INESC Coimbra, Coimbra, Portugal*
[b] *CeBER and Faculty of Economics, University of Coimbra, Portugal; lmcdias@fe.uc.pt*
[c] *Dept. of Electrical and Computer Engineering, University of Coimbra, Portugal; ch@deec.uc.pt*
[d] *GESEL, Federal University of Rio de Janeiro, Brazil; guilhermecrvg@yahoo.com.br, nivalde.castro@gmail.com*
[e] *EDP Brasil, Brazil, lucca.zamboni@edpbr.com.br*

Abstract: This work presents a methodology to perform a multi-criteria decision analysis based on qualitative assessments from multiple experts or stakeholders. The application that motivated these methodological developments is also described, consisting of an assessment of policies to foster technological innovations in the electricity sector, namely in a context of evolution to smart grids. Qualitative assessments were elicited from 28 experts using a Delphi process. The assessments concern the impact of policies on different objectives, but also opinions concerning the importance of these objectives. These opinions are translated into constraints on the weights of the corresponding criteria, in a partial information setting. Then, policies are sorted into categories based on a stochastic multi-criteria analysis for the ELECTRE TRI method. A new ranking approach is also developed to be compatible with the sorting results, based on a robustness analysis of ELECTRE TRI credibility indices. Results are discussed considering the entire set of experts and considering subsets of experts representing different perspectives.

Keywords: Multicriteria, Energy, Policy analysis, Delphi, Partial information, Outranking, MCDA

1. Introduction

Strategic decisions, including the assessment of business plans or public policies for possible implementation, entail several difficulties. Typically, stakes are high, decisions are costly to reverse, and consequences are endured throughout a long period. Furthermore, these decisions potentially affect different stakeholders whose support is important for the implementation of the plans or policies, and these stakeholders may have different views on the problem. Finally, the alternatives often need to be evaluated at a stage when they are not designed in detail, since the detailed specification of a policy or a plan can be a costly and time consuming process.

2

Several authors (e.g., [1–4]) sustained that Multi-Criteria Decision Analysis (MCDA) methods [5–7] are particularly well-suited to address strategic decision problems. MCDA provides a transparent way to encompass multiple potentially conflicting objectives, facilitating the incorporation of different stakeholder perspectives. MCDA based on multi-attribute value functions is often used to support policy making (e.g. [1,3,4,8,9]), but other MCDA approaches have also been used. ELECTRE methods [10,11], for instance, have been used in several applications (e.g., [12,13]) for their flexibility to deal with heterogeneous scales without transformations, to take into account imprecision in the input data, and mainly for not assuming full compensability among the criteria, i.e., excellent performance on one criterion may not compensate poor performance on another criterion [10].

The present work originated in a project included in the R&D Program regulated by ANEEL, the electricity sector Regulatory Authority in Brazil. Part of the project consisted of an assessment of policies to foster technological innovations in the electricity sector, in a context of evolution to smart grids. For such an assessment, all the difficulties mentioned above were present. A further complicating factor was the impossibility of gathering in a single workshop all the experts that would contribute to this assessment, ruling out a decision conference setting [8]. The participation of the experts involved in the assessment occurred through a Delphi survey process (Section 3.2), which was motivated by three main reasons. First, Delphi facilitated the elicitation of opinions from experts that were geographically dispersed. Secondly, it is a technique respondents were already familiar with, thus contributing to their engagement. Thirdly, it allowed asking elicitation questions in a qualitative way that did not assume expertise in MCDA.

The assessment consisted in sorting the policies into categories related with the implementation priority, from "Uninteresting" to "Implement with maximum priority", using the MCDA method ELECTRE TRI. This method evaluates each policy on its merits, and not in relation with other policies. It was also considered of interest to obtain a ranking of the alternatives that would be consistent with the ELECTRE TRI results.

The application-driven methodological developments presented in this paper aim to address two main challenges faced by the project team. One challenge was to obtain an ELECTRE TRI sorting of the policies based on Delphi assessments that used qualitative scales, not only to assess the impact of the policies on different criteria, but also to inquire about the importance of the criteria. According to a search on the Scopus® database only one relevant study that used an ELECTRE method with Delphi was found [14] ("electre" and "delphi" were the search keywords). The Co-oP decision support system offers several methods as independent components that include building ELECTRE outranking graphs, and Delphi and Nominal

Group Technique for group discussions [15]. However, none of these approaches use ELECTRE TRI or propose a way to set the criteria weights based on Delphi.

This work contributes to the literature by proposing a novel approach that translates Delphi qualitative assessments about the importance of the criteria into constraints on the weights that are exploited through a stochastic multi-criteria analysis (which is detailed in Sections 3 and 4). Another contribution is the proposal of a new reference-based ranking approach (inspired by the seminal work in [16]), in the form of a pessimistic and an optimistic variant. This approach builds on robustness analyses, in order to address the challenge of deriving a ranking of the policies in a way that is compatible with sorting results (also detailed in Sections 3 and 4). Finally, this work describes how a real-world application involving experts and decision makers in the Brazilian electricity sector motivated and exploited these methodological innovations.

The next section overviews the problem structuring process, describing the set of objectives, the policies to be assessed and the type of assessment sought. Section 3 presents the methodological contributions stemming from this work. The application of the methodology and the corresponding results are reported in Section 4. Section 5 concludes the paper by discussing the methodology and the results, as well as suggesting further research motivated by this work.

2. Problem structuring

Structuring the MCDA problem to be addressed is an essential phase in decision processes that involve multiple actors (thus multiple perspectives or even opposed interests), necessity of agreement (successful implementation depends on concerted action), or uncertainty over elements of the decision situation [17–19]. Problem structuring involved three main questions:

- What are the objectives of promoting technological innovation in the electricity sector, i.e. the objectives that justify the policies? In this work, each objective corresponds to an evaluation criterion that assesses the impact of each policy on the pursuit of that objective.

- Which policies are to be evaluated? Each policy corresponds to a potential action to be evaluated in the MCDA framework adopted.

- What type of assessment result is sought? Answering this question entails establishing what type of MCDA formulation (or *problématique*) is considered [20].

Pertaining to policies and underlying objectives, a literature review on experiences from 18 countries in implementing smart grid programs was conducted, and meetings were held with companies and regulatory authorities in Brazil, France, Italy, Germany, and Portugal. Research workshops were also organized. A structured process [21] using Soft Systems Methodology [17] and Value Focused Thinking [22] principles was followed to obtain a list of objectives:

O1 - *To benefit the environment and human health*, namely by increasing energy efficiency and the use of renewable sources, and avoiding construction of new plants and transmission lines. This objective includes greenhouse gas emissions, impact on human life and on other species, impact on other resources (land, water) and visual impact.

O2 - *To increase the flexibility and capabilities of the electricity system's technological infrastructure*. This includes real-time information on generation and consumption, ability to integrate distributed and intermittent renewable sources, remote management of equipment, ability to exploit distributed storage, ability to cope with the evolution of demand, reduction of technical and non-technical losses, and easier maintenance.

O3 - *To ensure security of supply*, in terms of energy independence from abroad, increasing reliability and quality of service, and taking into account vulnerability to attacks.

O4 - *To ensure openness, fairness, transparency and efficiency of the electricity markets*. This objective addresses openness to new producers, information to foster proactivity, access to energy services, equity among agents, efficient use of the available resources, encouragement of innovative business models, and electricity market risks.

O5 - *To provide financial benefit to the agents involved*. This objective encompasses investment and debt requirements, return on investment, and costs to the consumer (tariffs, energy efficiency).

O6 - *To provide economic and social benefit to the country*. This objective encompasses impact on the trade balance, impact on economic activity and employment and other impacts on the national budget, as well as training of human resources and technological leadership of the country.

O7 - *To ensure feasibility and to encourage adoption of technological innovations*. This objective addresses potential barriers to innovation, including capability of the agents, quality of telecommunications and technical support, privacy and security concerns, comfort, legal and bureaucratic barriers, and time required for implementation.

For a detailed presentation and justification of the set of objectives and the process to derive them the reader may consult [23].

The literature review, interviews and seminars that took place also informed the choice of policies to be evaluated, taking into account international experiences and the Brazilian reality:

P1 - *Mandatory roll out of smart meters*: this policy mandates distribution companies to install smart metering solutions at every consumer's location able to monitor energy flows in real-time, following the example of countries such as Italy. The costs of replacing existing meters by smart meters would be passed on to energy tariffs.

P2 - *Regulatory changes for technological innovation*: regulatory changes can foster technological modernization of electrical networks by recognizing investment in renewing and modernizing assets, and also by developing new methodologies to remunerate adequately technologies that entail a high ratio of operational expenditures (OPEX) to capital investments (CAPEX), following United Kingdom's RIIO (Revenue = Incentives + Innovation + Outputs) example.

P3 - *Improving R&D and demonstration project schemes*: this policy aims at increasing the participation of industry and the integration between projects. Emphasis is placed on larger projects and with higher Technology Readiness Level, following the example of Horizon 2020 in the EU.

P4 - *Incentives for demand side management, distributed generation and storage*: such incentives can take the form of differentiated feed-in tariffs for microgeneration (as occurs in several countries in the EU), tax incentives, and improved financing conditions.

P5 - *Definition of mandatory telecommunication quality requirements*: this policy addresses one of the main barriers to smart metering in Brazil, which is insufficient quality of telecommunication services.

P6 - *Regulatory changes for new business models*: regulatory changes are needed to facilitate the emergence of new business models that acknowledge new agents such as virtual power plants, energy aggregators as well as services based on energy efficiency and data analytics. In addition, distribution companies can be allowed to operate in non-regulated markets.

P7 - *Smart cities development plan*: this plan fosters investment in smart city programs encompassing multiple sectors (energy, telecommunications, transport, etc.), building on synergies among different services (e.g. electric mobility) and sharing of telecommunication and energy network infrastructures (and respective investment costs).

P8 - *National development plan of smart grid industries*: this plan consists of financial stimuli (to be decreased over time) for industrial development for a smart grids industry, complemented with the adoption of norms (namely for interoperability) allowing Brazilian industry to compete in international markets.

These policies are not mutually exclusive or exclusive of other alternatives. They were selected for being a rather diverse set of policies deemed relevant by the stakeholders and experts consulted.

Another question to be addressed at the structuring phase was the definition of the type of assessment result sought. The result should not be the selection of the best policy, because there was no reason why only one policy (or any other predefined number of policies) can be

6

implemented. A ranking would be insufficient, because a ranking gives information about how the alternatives compare to each other, but not on their intrinsic merits: it does not indicate if the best alternative in the ranking is good enough to be implemented.

The result should then be an ordinal sorting of the policies, partitioning the set of policies among ordered categories. Categories were defined a priori with a semantic meaning associated with implementation priority (this contrasts with clustering approaches in which categories are discovered a posteriori, e.g. [24]). A partition into four categories was deemed to be sufficient (from the worst to the best):

C1 - *Uninteresting*: the policy would lead to a situation that is equal to or worse than the current situation.

C2 - *Wait and see*: the policy would lead to a situation that is only slightly better than the current situation. There is no need to implement it at the moment.

C3 - *Implement with priority*: the policy would clearly contribute to improve the current situation.

C4 - *Implement with maximum priority*: the policy would clearly contribute to a large improvement to the current situation.

In addition to sorting the policies, it was also considered of interest to obtain a ranking of the policies within the same category by means of a process that would respect the sorting rationale, which is presented in Section 3. This ranking is a way to complement the sorting results (as mentioned above, a simple ranking was considered insufficient).

3. Evaluation methodology and process

3.1. MCDA aggregation model

The MCDA method used in this work needed to meet several requisites. It should be a sorting method, able to separate the policies into categories and such that the assessment of each policy is independent of other policies (as the model should be also applicable to other policies besides P1-P8). Secondly, it should not be a purely compensatory aggregation model in that a very good performance in one criterion is allowed to cancel out a very poor performance on another criterion. On the contrary, it should be possible to establish that a too negative performance on one criterion can put into question (veto) the implementation of a policy. Thirdly, it should be adequate to deal with qualitative judgments. Finally, it should be conceptually easy to communicate to an audience of stakeholders in the Brazilian electricity sector who are less familiar with MCDA.

ELECTRE TRI, a method firstly described by Yu [25] following the principles of ELECTRE [10] meets all the requirements outlined and has been widely applied [11]. It was built to deal

7

with quantitative evaluations, but it can be adapted to qualitative judgment. The analysts in this study were also experienced users of this method for problems in the energy sector (e.g., [13,26,27]). It was therefore a natural choice to be used in this work. Several workshops with stakeholders in the Brazilian electricity sector for discussing the methodology to be applied confirmed its adequacy. A summary of the ELECTRE TRI variant adopted and the parameters it involves is presented in Appendix A (for other variants, see e.g. [28–31]).

Each policy was described in about 200 words (the descriptions in Section 2 are just a summary). Indeed, a complete and accurate quantification of the impacts (economic environmental, etc.) would require an ambitious program of technical, economic, and environmental studies well beyond the scope of this project. Therefore, a more realistic option was followed: the policies were evaluated on a qualitative basis on each criterion, building on expert judgment from a diverse set of stakeholders. Table 1 presents the qualitative scale used for all the criteria and its relation with predefined profiles b^1 to b^3 that delimit the categories (Appendix A).

Table 1. Qualitative impact levels concerning the impact of each policy on each objective and their relation with predefined categories.

Level	Relation to category definitions
+5 extremely positive impact	Lower bound for C4 (b^3)
+4 very strong positive impact	
+3 strong positive impact	Lower bound for C3 (b^2)
+2 moderately positive impact	
+1 slight positive impact	Lower bound for C2 (b^1)
0 no impact	
-1 slight negative impact	
-2 moderately negative impact	
-3 strong negative impact	
-4 very strong negative impact	
-5 extremely negative impact	

The indifference and preference thresholds are the parameters that define how much significant a performance difference on a given criterion is. These are technical parameters that reflect the discriminating character of the criteria in face of imperfect knowledge, rather than preferences to be elicited [32]. It would be possible to set both parameters equal to zero. In such case, if the performance of a policy is below the lower profile of a category, then the criterion fully disagrees (concordance is null), otherwise the criterion fully agrees (concordance is 1).

Setting non-null indifference and preference thresholds allows considering a gradual transition from no concordance to full concordance. For this work the analysts proposed to use an indifference threshold $q_j=0.5$ and a preference threshold $p_j=2.5$ for all the criteria $j=1,...,7$, which was discussed and validated with a few key stakeholders. With these thresholds, the

concordance index (Appendix A) fully agrees that policy a_i outranks profile b^h when the performance of a_i is at least as good as the performance of b^h, it partially agrees if the performance of a_i is 1 or 2 levels below the performance of b^h (concordance of 0.75 and 0.25, respectively) and does not agree at all if the performance of a_i is 3 levels below the performance of b^h. These values were discussed at the same time that category profiles were defined, so that implications would be considered acceptable. For instance, to access category C3 "*Implement with priority*" (outranking b^2), a slight positive impact yields a modest concordance of 0.25, a moderately positive impact yields a large concordance of 0.75, and a strong positive impact yields a maximum concordance of 1.

These parameters were selected by the analysts taking into account the nature of the evaluation of policies in this project and the need to adapt ELECTRE TRI to qualitative evaluations. The remaining aspects of the model required judgment elicitation, as described below.

3.2. Delphi elicitation process

The Delphi process is a structured methodology to elicit opinions of multiple individuals who are subject to a series of questionnaires interspersed with controlled opinion feedback [33]. The inquired individuals do not engage in discussion among themselves, avoiding direct disputation.

Delphi was used in the present work to inquire about the opinion of different stakeholder perspectives on the impact of each policy on each objective, the importance of the objective, and the possibility of veto, which can be modelled in ELECTRE TRI (Appendix A). A pilot run of the questionnaire was first tried within the authors' organizations, allowing rewriting some questions in a clearer way before being sent to a larger group.

The group of experts that participated in the Delphi process consisted of 28 individuals, which could be loosely classified as representing three perspectives: a *Government perspective* (7 participants) including government and its agencies, state-owned companies, and government scientific advisors, a *Business perspective* (8 participants) including high-level representatives of private companies in the sector, and a *Knowledge perspective* (13 participants) including academics and consultants. Two rounds of the questionnaire took place, providing feedback to the participants from the first round to the second round.

Concerning the impact of the policies on the objectives, respondents answered the question: "Taking into account the objective of [*Name of the objective*] indicate what is the impact that you attribute to each one of the following policies". The answers were provided using the scale -5 to +5 presented in Table 1. Since this scale is an ordinal qualitative scale, the median, rather than an average, was used to summarize the answers of the respondents. This median was slightly adjusted by removing 0.5 for a few cases in which the number and levels of answers strictly below the median was much more relevant than the number and levels of answers

9

strictly above the median. For instance, the evaluation of policy P5 in Objective 1 had a median level +3 (chosen by 26% of the respondents), but 48% attributed a lower level (including a negative level for 4%), while 26% attributed a higher level (22% level +4, 4% level +5). In this case, the level most appropriate to represent the group's opinion was considered to be +2.5. This type of adjustment was made for 8 of the 56 evaluations.

Concerning the importance of the objectives, respondents answered the question: "Indicate your perspective on the importance of each one of the following objectives". The answers were provided using the following qualitative scale: 0 = of negligible importance, 1 = of little importance, 2 = moderately important, 3 = very important, 4 = strongly important, 5 = extremely important. This scale had the advantage of being easily understood by the respondents, but it does not allow a direct translation into weights without making the assumption that the scale is cardinal. This issue is addressed in a novel way, described in Section 3.3.

Concerning the veto power of the objectives, respondents answered the question: "Would you consider that if a policy has too much of a negative impact for *one* of the objectives then this would justify not implementing it, even if the impact on all other objectives is positive?" (yes/no). If yes then the following question was placed: "If you consider that a poor impact on an objective could in this way veto the implementation of a policy, indicate for each objective what impact levels you would consider negative enough to discard implementation. For objective [*Name of the objective*] a policy cannot be implemented now if the impact is equal to (or worse than) [*impact*]". The answers were provided using the scale presented in Table 1. For each objective, the median level among the answers was used to summarize the answers of the respondents. No other adjustment was deemed necessary.

The computation of ELECTRE's credibility indices (Appendix A) considers a binary veto effect similar to the ELECTRE TRI variant used by Öztürk et al. [34], which is compatible with the way the elicitation question was asked. For instance, if the highest level that implies veto to access category C3 is -2, then the veto threshold is set to 5 because -2 is 5 points below the level of the lower profile of C3, which is +3 (Table 1).

3.3 Translation of qualitative importance levels to weights

A direct translation of qualitative levels on the 0-5 scale would require assuming these levels were equally spaced (i.e. the difference between 0 and 1 is the same as the difference between 1 and 2, etc.). To avoid this assumption the solution adopted was to consider that the importance of an objective (a criterion) would depend on meta-weights associated with these levels, i.e., the weight of each criterion (denoted by k_j) should depend on:

$$I_j = \omega_0 p_{0j} + \omega_1 p_{1j} + \cdots + \omega_5 p_{5j} \ (j=1,...,7)$$

where ω_l ($l = 0, \ldots, 5$) denotes the meta-weight for level l and p_{lj} denotes the percentage of respondents responding level l for the importance of objective O_j. Thus, for any two objectives O_j and O_h, $I_j \geq I_h$ is interpreted as meaning $k_j \geq k_h$ in terms of criteria weights. This reasoning allows working with partial information without assuming precise values for the meta-weights, assuming only that the condition $\omega_0 \leq \omega_1 \leq \cdots \leq \omega_5$ should obviously be satisfied. According to Theorem 1 in [35], which was developed for the context of additive multi-attribute value functions, it holds that $I_j \geq I_h$, and consequently $k_j \geq k_h$, whenever the inequality $\sum_{i=l}^{5} p_{ij} \geq \sum_{i=l}^{5} p_{ih}$ holds for all $l = 0, \ldots, 5$. A partial ranking for the weights can thus be obtained; if neither $I_j \geq I_h$ nor $I_h \geq I_j$ hold for all $l = 0, \ldots, 5$, then in the absence of richer information no constraint is placed on the relation between k_j and k_h. Rather than choosing a single vector of weights, this work will consider all the weight values that respect the partial ranking derived as explained above.

3.4 Sorting and ranking approach

The assignment of each policy to a category is based on the concepts of robust and approximately robust conclusions [36,37]. Considering multiple equally admissible parameter vectors it is possible to determine robust classifications, i.e. cases in which a policy is always sorted into the same category for all admissible vectors [38]. If a policy is not robustly sorted into a single category, it is possible to determine the set of categories that are compatible with the admissible vectors.

This type of robustness analysis can be complemented with stochastic analyses, as suggested by [39] in a different context. A SMAA-TRI stochastic multi-criteria analysis [40,41] determines, for each policy, the proportion of the parameter vectors that lead to each possible category. This allows determining approximately robust classifications, i.e. cases in which a policy is almost always sorted into the same category. In this work, a classification was deemed to be (approximately) robust if it held for 99.9% or more of all admissible vectors.

Admissible parameter vectors were defined by three conditions: weights should respect the partial order described in Section 3.3; the minimum admissible weight is 0.085, so that no criterion should weight more than the six remaining criteria altogether (the implicit maximum weight is 0.49). It would have been possible to simply set the constraints $k_j \leq 0.49$, but the strategy followed not only ensures this inequality, it also ensures no criterion has null weight, which was considered a desirable feature.

Bounds for the cutting level λ (Appendix A) were discussed at an informal academic meeting where some stakeholders were present to analyze preliminary results. Demanding unanimity, (i.e. all seven criteria in favor, with $\lambda = 7/7$) was deemed unnecessarily stringent. The discussion led to constrain the cutting level to lie between 4/7 and 5/7. The lower bound corresponds to the

support of a coalition of criteria that represents over 57.1% of the criteria weights. For a vector of equal weights, this requires the support of 4 out of the seven criteria, although the criteria weight bounds allow this level to be achieved by two single criteria. The upper bound corresponds to the support of a coalition of criteria that represents over 71.4% of the criteria weights, which was deemed to be a comfortable majority. Since these limits are different from the usually "round" limits set by decision makers (e.g., 1/2, 2/3, 3/4), results are also presented considering an interval $\lambda \in [1/2, 3/4]$.

In the preliminary presentation of the sorting results, it was considered that stakeholders would also welcome a ranking of the policies. This raised an interesting question, since a ranking procedure such as ELECTRE III [10] would not be appropriate. Indeed, such procedures would compare the policies against each other to obtain a partial ranking that might contradict the sorting results. Moreover, adding or deleting a policy might change the relative position of the remaining policies. The solution would need to be a reference-based ranking approach, based on sorting profiles, a type of approach that has been proposed and studied only in recent years [16,42].

In a situation of partial information, where multiple parameter vectors are admissible, it is necessary that a reference-based ranking of the policies is compatible with robust sorting conclusions. For this purpose, a reasonable and easy to communicate solution is to base the ranking of the policies within each sorting category on the strength of the arguments supporting the attainment of the proposed classification. In ELECTRE TRI, this is given by the credibility that the policy under analysis outranks the category's lower bound (Appendix A). This credibility varies with the weight vector, but it is possible to solve a linear program to obtain the minimum and the maximum credibility subject to the constraints imposed on the weights [43].

Let $s^M(a_i, b^h)$ and $s_m(a_i, b^h)$ denote respectively the maximum and minimum overall credibility of the assertion that policy a_i outranks b^h, considering as variables the weights and the cutting level. These values can be obtained by maximizing and minimizing $c(a_i, b^h)$ using linear programming formulations (for detailed formulations, see [43]).

Let $W(a_i)$ and $B(a_i)$, respectively, denote the indices of the worst and best category where policy a_i can be sorted when satisfying the constraints placed on the parameter values ($i=1,\dots,m$). If the interval of variation for the cutting level is $\lambda \in [\lambda_{min}, \lambda_{max}]$, then $W(a_i)$ is the highest value $h \in \{1,\dots,k\}$ such that $s_m(a_i, b^{h-1}) \geq \lambda_{max}$. On the other hand, $B(a_i)$ is the lowest value $h \in \{1,\dots,k\}$ such that $s^M(a_i, b^h) < \lambda_{min}$. Let us note that in some circumstances there may be categories in-between $W(a_i)$ and $B(a_i)$ where a_i cannot be sorted [43].

To obtain a ranking of the policies this work followed a pessimistic (or cautionary) perspective taking into account the principles outlined next:

Pr1) Policies are ranked according to the worst category they can have: if $W(a_i)>W(a_j)$, then a_j should have a worse rank. As a particular case, policies robustly sorted into C^h (i.e., with $W(a_i)=B(a_i)=h$) are better ranked than those robustly sorted into a worse category C^q ($q<h$).

Pr2) If two policies a_i and a_j have the same worst category $W(a_i)=W(a_j)$, but a_i can potentially reach a better category than a_j, i.e. $B(a_i)>B(a_j)$, then a_j should have a worse rank.

Pr3) If two policies a_i and a_j have the same worst category $W(a_i)=W(a_j)=\omega$ and they also have the same best category $B(a_i)=B(a_j)=\beta$, a lexicographic comparison of vectors $(s_m(a_i,b^{\omega-1}), \dots, s_m(a_i,b^{\beta}))$ and $(s_m(a_j,b^{\omega-1}), \dots, s_m(a_j,b^{\beta}))$ is used to obtain a more discriminating ranking, with the rationale that the better ranked alternative would be last one to worsen its worst category if λ_{max} increases:

a_i is ranked better than $a_j \Leftrightarrow s_m(a_i,b^{\omega-1})>s_m(a_j,b^{\omega-1})] \vee$

$$\exists g \in [\omega,\beta]: s_m(a_i,b^h)=s_m(a_j,b^h) \text{ for all } h=\omega-1,\dots,g-1 \wedge s_m(a_i,b^g)>s_m(a_j,b^g)].$$

It is debatable whether a policy with $W(a_i)=B(a_i)=3$ should be better ranked than a policy with a $W(a_j)=2$ and $B(a_j)=4$. This rule focuses on the worst case, which adheres to the "pessimistic" or "conservative" character of the ELECTRE TRI variant used in this work. It would be equally possible to follow an optimistic or "benefit of doubt" procedure as follows:

Pr1') Policies are ranked according to the best category they can have: if $B(a_i)>B(a_j)$, then a_j should have a worse rank.

Pr2') If two policies a_i and a_j have the same best category $B(a_i)=B(a_j)$, but a_i can potentially be in a worse category than a_j, i.e. $W(a_i)<W(a_j)$, then a_j should have a better rank.

Pr3') If two policies a_i and a_j have the same worst category $W(a_i)=W(a_j)=\omega$ and they also have the same best category $B(a_i)=B(a_j)=\beta$, a lexicographic comparison of vectors $(s^M(a_i,b^{\beta}),\dots, s^M(a_i,b^{\omega-1}))$ and $(s^M(a_j,b^{\beta}),\dots, s^M(a_j,b^{\omega-1}))$ is used, since the best category is defined by a maximum credibility.

Both procedures implement the basic principle that if $W(a_i)\geq W(a_j)$ and $B(a_i)\geq B(a_j)$, and one of this inequalities is strict, then a_i is better ranked than a_j.

4. Results

4.1. Results considering all respondents

This subsection reports the results considering the answers of all 28 respondents who participated in the two rounds of the Delphi process described in Section 3. Concerning the impact of the policies on the different objectives the results are presented in Table 2. Referring to the descriptions in Table 1, respondents consider that all policies have a positive impact for

the different objectives, and it is clear that no policy will be sorted into category C^1 (*Uninteresting*). This corresponds to the expectation of the project team had when conceiving the policies to be evaluated.

Table 2. Performance table obtained from the Delphi process (all respondents).

	O1 Environ. & human health	O2 Technol. infra-structure	O3 Security of supply	O4 Electricity markets	O5 Financial benefit to agents	O6 Benefit to country	O7 Feasibility & adoption
P1 - Roll out smart meters	1.5	3.5	2	2.5	2.5	2	2
P2 - Regulatory changes	3	4	3.5	4	3	4	4.5
P3 - R&D and demonst.	3	4	3	3	3	4	4
P4 - DSM/DG/S incentives	4	4	4	4	3.5	4	4
P5 - Telecom standards	2.5	3	2	2	1.5	3	2.5
P6 - New business models	3	4	3	4	4	4	4
P7 - Smart cities	4	4	3	3	3	4	4
P8 - Smart grid industries	3	4	3	3	3	4	4

Since the profiles defining the categories and the indifference and preference thresholds have been fixed (Section 3.1), it is now possible to compute single-criterion concordance indices, as described in Appendix A. These indices indicate how much each criterion agrees with an outranking relation, from a minimum of 0 to a maximum of 1. The evaluations provided by the respondents are sufficient, with two exceptions (Policy P1 on objective O1 and policy P5 on objective O5), to guarantee a strong support for a policy to be at least sorted into category C^3, with concordance 0.75 or higher (Table B1 in Appendix B). Concordance indices for attaining category C^4 are obviously lower, but still strong in many cases (Table B2 in Appendix B).

Concerning discordance, respondents were asked to indicate what impact levels they would consider negative enough to discard implementation of a policy. According to the median responses in the second round of the Delphi process, this level would be -1.5 for the first criterion (*Environ. & human health*) and it would be -2 for the remaining criteria. Accordingly, as explained in Section 3.2, the veto threshold for the first criterion was set to 4.5 (difference from -1.5 to level 3) and the veto threshold for the remaining criteria was set to 5 (=3-(-2)). These veto thresholds are such that a veto situation only occurs if at least half of the respondents would agree. The single-criterion discordance indices (Appendix A) indicate whether a criterion vetoes an outranking relation. With these inputs, the discordance indices for attaining C^3 and C^4 are all null (no veto situation occurs).

Given these single-criterion concordance and discordance indices, outranking relations and the corresponding sorting of the policies depend solely on the criteria weights and the value of the

14

cutting level. As explained in Section 3, respondents provided their opinion on the relative importance of each objective (criterion) using a qualitative scale ranging from 0 (negligible importance) to 5 (extremely important). Table 3 summarizes the responses after the second round of the Delphi process.

Table 3. Relative importance of the criteria (percentage of all respondents choosing each level).

Level	O1 (g_1) Environ. & human health	O2 (g_2) Technol. infra- structure	O3 (g_3) Security of supply	O4 (g_4) Electricity markets	O5 (g_5) Financial benefit to agents	O6 (g_6) Benefit to country	O7 (g_7) Feasibility & adoption
0	0%	0%	0%	0%	0%	0%	0%
1	4%	0%	4%	0%	0%	4%	0%
2	14%	0%	0%	0%	4%	0%	4%
3	32%	18%	11%	25%	36%	0%	14%
4	25%	50%	46%	46%	43%	25%	46%
5	25%	32%	39%	29%	18%	71%	36%

There are two strategies in MCDA for deriving results that take into account the different perspectives concerning the importance of the criteria [44]: one is to obtain results for each individual by running the MCDA method from start to finish using his/her inputs and then aggregate these individual outputs (e.g. [45,46]), another one is to aggregate preferences at the inputs level and then obtain results for the collective input. An example of the former strategy has been proposed by Damart et al. [45] for ELECTRE TRI, in which each individual indicates the category assignments that would be acceptable for him/her based on his/her information. To obtain such results for each respondent, we considered a variation interval of [4/7, 5/7] for the cutting level and a set of constraints on weights that respected the ordinal importance levels defined by that individual: if a criterion has a higher importance than another one, then the first one should have more weight. If two criteria have the same level, they were not constrained to have the same weight. Otherwise, since there were six different levels (from 0 to +5) to be assigned to seven criteria, at least two of them would be constrained to have the same weight. As an example, a given respondent stated that O1 and O5 were very important (level 3), O3, O4, O6 and O7 were strongly important (level 4), and O2 was extremely important (level 5). This originated the following constraints:

$$k_2 \geq k_j + \varepsilon, \, k_j \geq k_1 + \varepsilon, \text{ and } k_j \geq k_5 + \varepsilon \, (j = 3,4,6,7).$$

The value of ε was arbitrarily set to 0.01 to enforce a strict inequality. A mathematical program approach was then used to maximize and minimize $s(a_i, b^h)$ for different category profiles (for details, see [47]). The results are presented in Table 4. Except P4, all policies could be sorted into C3 with unanimity (P4 is also acceptable in that category for 61% of the respondents), and

15

it is also unanimous that P2, P4 and P6 could be sorted into C4 (P7 is also acceptable in that category for 71% of the respondents). These results yield a clear picture of what classifications are not supported by the group, but do not offer much discrimination between the policies (Damart et al. [45] proposed an iterative process of multiple discussion rounds to obtain a more clear-cut result).

Table 4. Proportion of respondents that would accept each potential classification considering common cutting level bounds $\lambda \in [4/7,5/7]$ and their individual weight constraints

	C1 Uninteresting	C2 Wait and see	C3 Implement with priority	C4 Implement with maximum priority
P1 - Roll out smart meters			100%	
P2 - Regulatory changes			100%	100%
P3 - R&D and demonstration			100%	
P4 - DSM/DG/S incentives			61%	100%
P5 - Telecom standards			100%	
P6 - New business models			100%	100%
P7 - Smart cities			100%	71%
P8 - Smart grid industries			100%	

An alternative process of aggregating information at the inputs level was followed in this work. Using the average of the responses to set the criteria weights would not be legitimate because of the qualitative nature of the scale used. Using the median of the responses would be more defensible, but it would still require a questionable operation to transform these medians into a vector of weights that would later be added. It is more appropriate in this case to acknowledge the variety of opinions about these crucial parameters and to consider a set of multiple admissible weight vectors respecting Table 3.

A strict consensus about the ranking (or a partial ranking) of the weights did not arise from the responses, i.e., there are no two criteria g_j and g_k such that all the respondents assigned a level of importance to g_j which was higher than the level they assigned to g_k. However, if one compares, for instance, criteria g_4 and g_5 it is possible to notice the following: 100% of the respondents agree that g_4 is at least level 2 (and the same applies to g_5); 100% of the respondents agree that g_4 is at least level 3 (but only 96% of them for g_5); 75% of the respondents agree that g_4 is at least level 4 (but only 61% of them for g_5); and finally, 29% of the respondents agree that g_4 is level 5 (vs. 18% for g_5). As explained in section 3.3, this provides a compelling rationale to require $k_4 > k_5$. Similar conclusions can be made concerning other pairs of criteria, based on the cumulative percentage of responses supporting that the criteria importance level is at least l (for $l = 0,\dots,5$) (Figure 1). Figure 2a) depicts the resulting partial ranking of the criteria weights.

16

A few additional weaker inequalities were added, to reflect other pairs for which an examination of the individual responses (not presented here) indicated there was a large agreement: $k_6 \geq k_2$ (82% of the respondents agreed) and $k_3 \geq k_5$ (89% agreed), implying by transitivity $k_6 \geq k_4$ (96% agreed) and $k_6 \geq k_5$ (93% agreed). The resulting relation is depicted in Figure 2b). Contrarily to the relations in Figure 2a), which were considered as strict inequalities (a minimum arbitrarily small difference of 0.005 was required), $k_6 \geq k_2$ and $k_3 \geq k_5$ were considered as weak inequalities (admitting equality). Finally, to prevent a criterion from having a null weight, a minimum weight was required. Since in this case the minimum weight cannot be set based on limits to trade-offs (as in, e.g. [48]), the lowest minimum considered reasonable was chosen to be 0.085, so that no criterion would have null weight or a weight higher than $1 - 6 \times 0.085 = 0.49$.
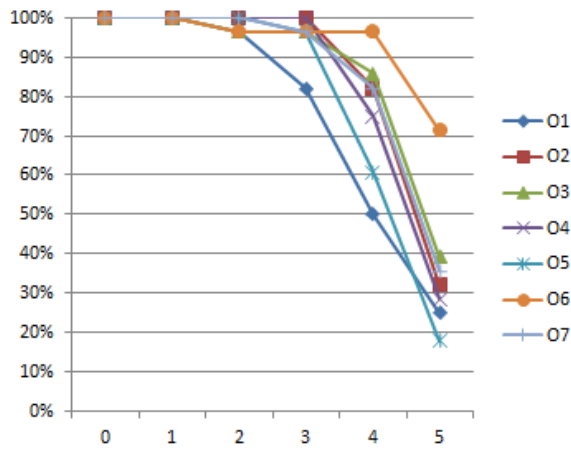

Figure 1. Cumulative percentage of responses supporting that the criteria importance level is at least l (for l = 0,...,5).
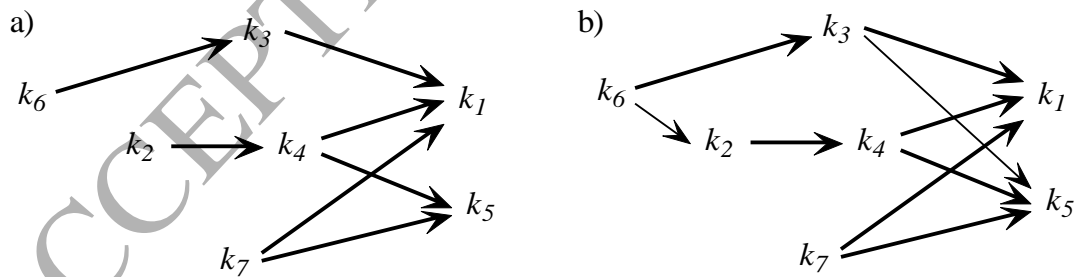

Figure 2. Partial ranking of the criteria weights (all respondents): a) based on cumulative responses; b) based on cumulative responses and weaker relations.

A stochastic multi-criteria analysis was conducted considering the constraints on the weights and a variation interval of [4/7, 5/7] for the cutting level. A hit-and-run Monte-Carlo procedure [49] considering 50000 draws from uniform distributions was used for this purpose. Results are presented in Table 5. Four policies are robustly sorted into C^3 (*Implement with priority*) and one

policy is robustly sorted into C^4 (*Implement with maximum priority*). The remaining three policies display a robustness interval $[C^3, C^4]$, meaning they are sorted into C^3 for some parameter vectors and they are sorted into C^4 for other parameter vectors.

Table 5. Sorting probabilities considering weight constraints and cutting level bounds $\lambda \in [4/7, 5/7]$ (all respondents). In parenthesis: probabilities considering $\lambda \in [1/2, 3/4]$, if different.

	C1 Uninteresting	C2 Wait and see	C3 Implement with priority	C4 Implement with maximum priority
P1 - Roll out smart meters	0	0	1.000	0
P2 - Regulatory changes	0	0	0.391 *(0.369)*	0.609 *(0.631)*
P3 - R&D and demonstration	0	0	1.000 *(0.916)*	0 *(0.084)*
P4 - DSM/DG/S incentives	0	0	0 *(0.099)*	1.000 *(0.901)*
P5 - Telecom standards	0	0	1.000	0
P6 - New business models	0	0	0.567 *(0.471)*	0.433 *(0.529)*
P7 - Smart cities	0	0	0.962 *(0.727)*	0.038 *(0.273)*
P8 - Smart grid industries	0	0	1.000 *(0.916)*	0 *(0.084)*

Table 6. Sorting based on the stochastic analysis (all respondents).

	Sorting (based on stochastic analysis)
P1 - Roll out smart meters	Implement with priority
P2 - Regulatory changes	Implement with high priority[*]
P3 - R&D and demonstration	Implement with priority
P4 - DSM/DG/S incentives	Implement with maximum priority
P5 - Telecom standards	Implement with priority
P6 - New business models	Implement with high priority[*]
P7 - Smart cities	Implement with high priority[*]
P8 - Smart grid industries	Implement with priority

[*] Sorting varies between "Implement with priority" and "Implement with maximum priority"

Table 6 presents the sorting results in a way that facilitates communication with stakeholders making use of a label *Implement with high priority* to represent the interval $[C^3, C^4]$, i.e. hesitation between C^3 and C^4. Indeed, policies in the interval $[C^3, C^4]$ can be considered as having more priority than policies in C^3 and having less priority than policies in C^4. Results were also obtained considering a wider variation interval of $[1/2, 3/4]$ for the cutting level, to

assess the extent to which conclusions would change (Table 5, values in parenthesis). By accepting a more demanding value for the cutting level, 0.75, lower classifications become possible, and this occurs only for P4 in 9.9% of the cases. On the other hand, by accepting a less demanding value for the cutting level, 0.50, higher classifications become possible, and this occurs only for P3 and P8 in 8.4% of the cases. The remaining minimum and maximum classifications would not change.

Finally, a ranking of the policies within each category or interval of categories was obtained as described in Section 3.4. Figure 3 depicts the credibility that a policy may reach category C^3 (i.e. credibility of outranking profile b^2), whereas Figure 4 depicts the credibility that a policy may reach category C^4 (i.e. credibility of outranking profile b^3). The minimum and maximum credibility were determined using linear programs minimizing or maximizing concordance (similar to [43]), subject to the constraints on the weights (partial ranking and lower bound). The same results are presented as a table in Appendix B (Table B3).

P4 is the only policy robustly sorted into C^4 and is ranked at the top. Next, policies that can be sorted into C^3 or C^4 are considered: P2, P6 and P7. Comparing lexicographically the minimum credibility of outranking b^2, b^3 and b^4 (b^4 is a theoretical maximum that cannot be outranked, hence we set $s_m(.,b^4)=0$ for all policies), P2, P6 and P7 (starting from the best one) enter the ranked list. Finally, policies robustly sorted into C^3 are considered, comparing lexicographically the minimum credibility of outranking b^2 and b^3: first P3 and P8 (a tie that cannot be broken), then P5, followed by P1 (Table 7). As a curiosity, the same ranking would result if the optimistic or "benefit of doubt" procedure Pr1'-Pr3' was followed.
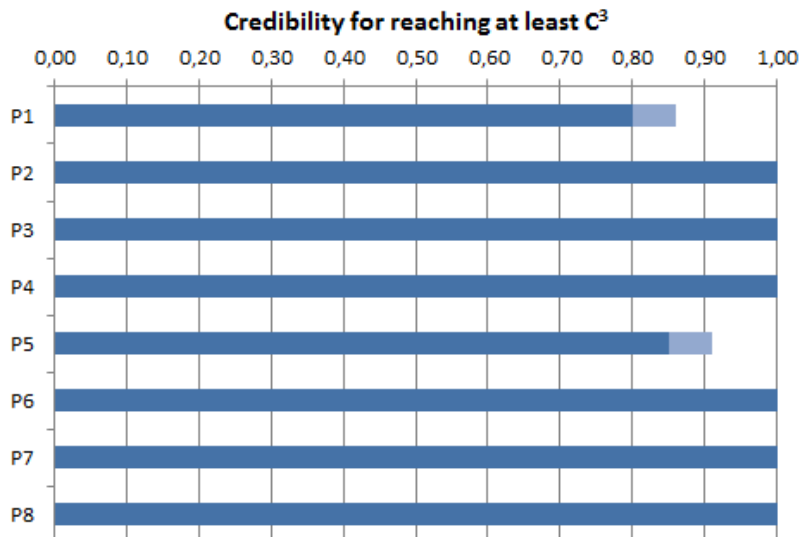

Figure 3. Robustness analysis: minimum and maximum credibility of outranking b^2, i.e. the lower bound of C^3, considering weight constraints (all respondents).
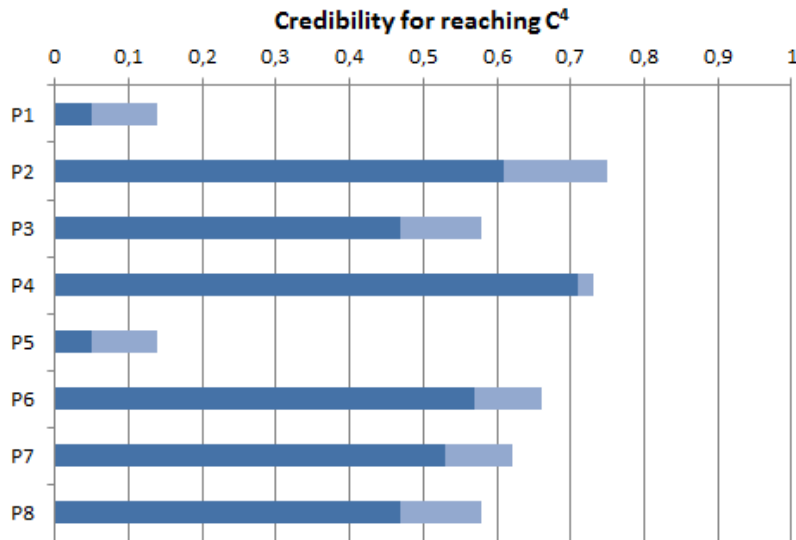
Figure 4. Robustness analysis: minimum and maximum credibility of outranking b^3, i.e. the lower bound of C^4, considering weight constraints (all respondents).

Table 7. Ranking based on the robustness analysis (all respondents).

	W(.)	B(.)	Vector for lexicographic tie break	Ranking
P4 - DSM/DG/S incentives	4	4	*(no need for a tie break)*	1st
P2 - Regulatory changes	3	4	$(s_m(.,b^2), s_m(.,b^3), s_m(.,b^4))=(1, 0.61,0)$	2nd
P6 - New business models	3	4	$(s_m(.,b^2), s_m(.,b^3), s_m(.,b^4))=(1, 0.57,0)$	3rd
P7 - Smart cities	3	4	$(s_m(.,b^2), s_m(.,b^3), s_m(.,b^4))=(1, 0.53,0)$	4th
P3 - R&D and demonstration	3	3	$(s_m(.,b^2), s_m(.,b^3))=(1, 0.47)$	5/6th
P8 - Smart grid industries	3	3	$(s_m(.,b^2), s_m(.,b^3))=(1, 0.47)$	5/6th
P5 - Telecom standards	3	3	$(s_m(.,b^2), s_m(.,b^3))=(0.85, 0.05)$	7th
P1 - Roll out smart meters	3	3	$(s_m(.,b^2), s_m(.,b^3))=(0.80, 0.05)$	8th

4.2. Results considering different perspectives

The analysis reported in Section 4.1 was repeated considering subsets of the responses to the Delphi survey. One subset of 7 responses from government-related officials and state companies represents a *Government perspective*. A second subset of 8 responses from managers in private companies represents a *Business perspective*. A third subset of 13 responses from academics and consultants represents a *Knowledge perspective*. This section presents the main results obtained for these three perspectives. Inputs and intermediate results are presented in Appendices C, D and E (supplementary material).

Tables C1, D1 and E1 in Appendices C, D and E indicate the responses considered to be the most representative for each perspective. In most cases these were exactly equal or 0.5 points away from the overall representative levels used in Section 4.1. The exceptions are highlighted in Table 8, displaying the evaluation levels where some disagreement can be found.

Table 8. Performance table obtained from the Delphi process (different perspectives).

	O1 Environ. & human health	O2 Technol. infra-structure	O3 Security of supply	O4 Electricity markets	O5 Financial benefit to agents	O6 Benefit to country	O7 Feasibility & adoption
P1 - Roll out smart meters	≈	≈	=	B↑	G↓ B↑	≈	≈
P2 - Regulatory changes	≈	=	≈	=	=	≈	≈
P3 - R&D and demonst.	≈	B↓	≈	≈	B↓	≈	≈
P4 - DSM/DG/S incentives	≈	=	G↑	=	=	G↑	=
P5 - Telecom standards	B↓	≈	B↓	≈	≈	≈	G↑
P6 - New business models	B↓	=	≈	≈	=	B↓	B↑
P7 - Smart cities	≈	G↓	≈	≈	≈	=	≈
P8 - Smart grid industries	G↑	≈	B↑	≈	≈	=	=

Legend: = denotes all perspectives indicated the same level; ≈ denotes a difference of at most ±0.5 relatively to Table 2; B↑ (B↓) denotes the *Business perspective* indicated a better (resp. worse) level by a difference of 1 or more relatively to Table 2; G↑ (G↓) denotes the same but for the *Government perspective*.

Following the same methodology used for the global set of responses (Section 4.1), veto thresholds were computed and partial rankings of criteria weights were derived. Table 9 allows comparing the veto thresholds for each perspective, showing that the *Government perspective* is the most demanding about minimum requirements for implementation (lower veto thresholds mean higher veto capacity). Tables C2-C4, D2-D4 and E2-E4 in Appendices C, D and E indicate the concordance and discordance indices for attaining C^3 and for attaining C^4. The discordance indices for attaining C^3 are not displayed, as they were always null.

Tables C5, D5 and E5 in Appendices C, D and E summarize the Delphi responses concerning the importance of each objective, from which the partial rankings of criteria weights were derived. The only contradictory opinion found in these relations (Figure 5) concerns the importance of the second and the fourth criteria. From the *Government perspective* $k_4 > k_2$, whereas the *Business perspective* and the *Knowledge perspective* display $k_2 > k_4$.

Table 9. Veto thresholds for the different perspectives.

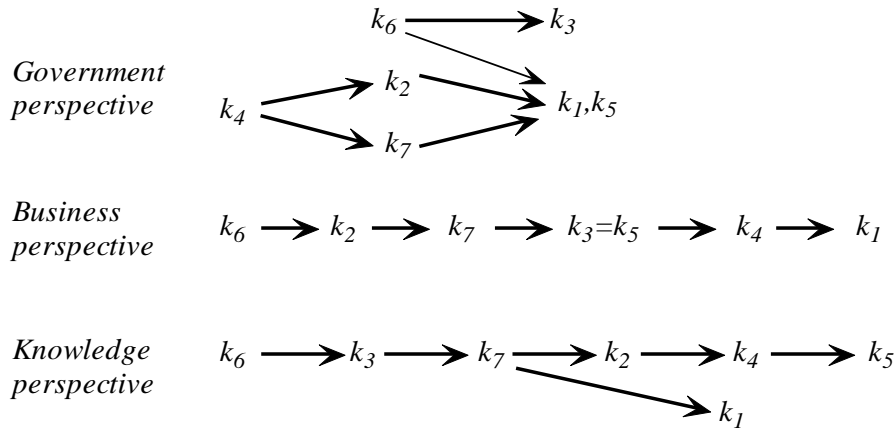| Perspective	v_1	v_2	v_3	v_4	v_5	v_6	v_7
Government	4	5	4	5	5	4	5
Business	4	6	5	6	5	5	6
Knowledge	5	5	5	5	5	5	5

21

Figure 5. Partial ranking of the criteria weights from different perspectives based on cumulative responses. All relations are strict (>) except $k_6 \geq k_1$ and $k_6 \geq k_5$ for the Government perspective.

It is noteworthy and unexpected that the first criterion (Environ. & human health) tends to be considered one of the least important in terms of its weight, but it is the criterion with the greatest veto power (lowest veto threshold). It appears that having at worst a slightly negative impact on this criterion is a necessary condition for implementation, but once this minimum performance is met the criterion does not matter as much as the remaining ones.

Given these constraints on the weights, the sorting and ranking of the policies for each perspective can be computed. Table 10 presents the sorting results, which are based on the sorting probabilities presented in Tables C6, D6 and E6 in Appendices C, D and E. Table 11 presents the ranking results, which are based on the minimum and maximum credibility indices presented in Tables C7, D7 and E7 in Appendices C, D and E.

Table 10. Sorting based on the stochastic analysis (different perspectives).

Policy	Government perspective	Business perspective	Knowledge perspective
P1 - Roll out smart meters	C^3: priority	C^3: priority	C^3: priority
P2 - Regulatory changes	$[C^3, C^4]$: *high priority*	$[C^3, C^4]$: *high priority*	$[C^3, C^4]$: *high priority*
P3 - R&D and demonstration	C^3: priority	C^3: priority	C^3: priority
P4 - DSM/DG/S incentives	C^4: max. priority	C^4: max. priority	$[C^3, C^4]$: *high priority*
P5 - Telecom standards	C^3: priority	$[C^2, C^3]$: *no priority*	C^3: priority
P6 - New business models	C^4: max. priority	C^3: priority	C^3: priority
P7 - Smart cities	C^3: priority	$[C^3, C^4]$: *high priority*	$[C^3, C^4]$: *high priority*
P8 - Smart grid industries	$[C^3, C^4]$: *high priority*	$[C^3, C^4]$: *high priority*	C^3: priority

Table 11. Ranking based on the robustness analysis (different perspectives).

Rank	Government perspective	Business perspective	Knowledge perspective
1st	P4 -DSM/DG/S incentives	P4 -DSM/DG/S incentives	P4 -DSM/DG/S incentives
2nd	P6 - New business models	P7 - Smart cities(*)	P2 - Regulatory changes(*)
3rd	P2 - Regulatory changes	P8 - Smart grid industries(*)	P7 - Smart cities(*)
4th	P8 - Smart grid industries	P2 - Regulatory changes	P6 - New business models
5th	P7 - Smart cities	P6 - New business models	P3 - R&D and demonstration
6th	P3 - R&D and demonstration	P3 - R&D and demonstration	P8 - Smart grid industries
7th	P5 - Telecom standards	P1 - Roll out smart meters	P1 - Roll out smart meters
8th	P1 - Roll out smart meters	P5 - Telecom standards	P5 - Telecom standards

(*) *ex-aequo*

Overall, the sorting for each perspective is not very different from the sorting obtained considering the entire set of responses (Table 6). According to all the perspectives analyzed, all policies have merits for implementation with some degree of priority. The only exception is policy P5 under the *Business perspective*, where it could belong to category C^2 (Wait and see) for some of the parameter values accepted. This policy, together with P1, is at the bottom of the ranking for the three perspectives. All perspectives rank policy P4 at the top. Comparing the different perspectives, results are in agreement to a large extent since there is always a category in common for every policy, except P6.

5. Discussion and conclusions

The methodology proposed in this work was motivated by a real-world application and its application demonstrated its usefulness. The problem consisted of evaluating different policies that were not developed in sufficient detail to allow measuring its impacts in a quantitative manner.

The methodology followed in this work is based on a well-known and proven multi-criteria sorting method: ELECTRE TRI (pseudo-conjunctive variant). Expert judgment from 28 stakeholders was obtained by means of a Delphi survey eliciting inputs on a qualitative scale about policy impacts and some of the main parameters of the ELECTRE TRI model: criteria weights and veto thresholds. However, the methodology needed to be developed to meet the specific requirements of this application.

First, it was necessary to define category profiles, indifference and preference thresholds respecting the ordinal nature of the evaluation scales. This was possible using Table 1 as a means of communicating with stakeholders. A similar reasoning was necessary pertaining to veto thresholds. The elicitation of these parameters was based on the critical transition from category C^2 (Wait and see) to C^3 (Implement with priority).

A more relevant methodological contribution is the process suggested to sort and rank the policies. Sorting is based on a SMAA-TRI type of stochastic analysis based on a partial ranking of the criteria weights derived from the hypothesis of weighting (but not assuming any specific meta-weights) the qualitative importance levels provided by the experts. The proposed ranking methodology is based on a robustness analysis of sorting results, thereby ensuring that a policy is always ranked in a better position than another policy sorted into a worse category or interval of categories (assuming a worst-case evaluation as a matter of caution).

The analysis was conducted for the entire set of responses, but also for subsets representing different perspectives. Results obtained for each perspective highlight where disagreement might exist. For instance, the sorting results considering all the responses put policies P2 (Regulatory changes) and P6 (New business models) into the same robustness interval $[C^3, C^4]$ (Table 6). But perspective-wise results in Table 10 unveil an important difference. Policy P2 is sorted into the robustness interval $[C^3, C^4]$ for all the perspectives, whereas Policy P6 is robustly sorted into C^4 by the *Government perspective* and it is robustly sorted into C^3 by the *Business* and *Knowledge perspectives*. This hints that if Government seeks to implement this policy with maximum priority, then it should look into the concerns of the other perspectives when designing the policy in detail.

The methodology and the results have been publicly presented to audiences of stakeholders in academia (open lecture at Federal University of Rio de Janeiro), electricity sector representatives (EDP Brasil, S. Paulo, with other electricity companies also present) and at the electricity sector Regulatory Authority (ANEEL, Brasilia). Acceptance of the methodology and results has been quite encouraging in all presentations, suggesting the continuation of this line of research. A few potentially interesting questions for future research are outlined next.

A limitation of this work is that it does not address potential synergies among policies, as policies were evaluated independently of each other considering their intrinsic advantages and disadvantages. If it was necessary to consider synergies between a few policies only, the same methodology could be used to evaluate composite policies (e.g. a policy P9 could represent implementing two other policies at the same time), but evaluating all combinations might be impractical. Future research might however assess synergies between the most promising policies (according to the outputs of this study). An interesting extension of this methodology would be to view the problem as a portfolio evaluation one, where each portfolio would consist of multiple policies.

The research conducted in this work may also be repeated in different contexts, or even in the same context but with different stakeholders. Indeed, the results are contingent on the views of the experts who collaborated in this work. The larger the number and the variety of stakeholders involved, the more representative results will be. In one of the public presentations of this work

24

it was suggested that a fourth perspective could be that of the general public (who are mainly consumers). Consumer representatives might not necessarily be experts capable of assessing the impacts of policies on the different objectives, but they could express their views on the questions concerning the importance of the criteria and the possibility of veto. Future applications of this methodology can then use different sets of respondents to answer different questions in the Delphi survey, thus allowing the elicitation of the views of non-experts concerning aspects that do not require domain-specific expertise.

Furthermore, in future applications, different criteria might be assessed by different specialized experts, possibly at a sub-criteria level, in which each objective would be divided into sub-objectives to be evaluated separately. This will require adapting the methodology proposed in this paper to a context of a Multiple Criteria Hierarchy Process [50,51].

Appendix A. ELECTRE TRI computations

ELECTRE TRI is an approach that admits several variants. The variant of ELECTRE TRI used in this work is the one proposed by Mousseau and Dias [52,53]. The inputs are the following:

- $A=\{a_1,\dots,a_m\}$ is the set of m alternatives to be sorted.

- $G=\{g_1(.),\dots,g_n(.)\}$ is a set of n evaluation criteria. The following computations assume that (as is the case in this work), the higher is $g_j(a_i)$, the better is the performance of alternative a_i on criterion $g_j(.)$.

- $B=\{b^0,\dots,b^k\}$ denotes the set of profiles that define a set of k categories, $\{C^1,\dots,C^k\}$. Each category C^h ($h=1,\dots,k$) is defined through two reference profiles: a lower bound b^{h-1} and an upper bound b^h. By convention, C^1 denotes the worst category and C^k denotes the best category.

For each criterion ($j=1,\dots,n$) the following preference-related parameters are set:

- q_j denotes the indifference threshold for criterion $g_j(.)$: two levels of performance are considered to be indistinguishable if their difference is q_j or less.

- p_j denotes the preference threshold for criterion $g_j(.)$: one level of performance is considered to be undoubtedly better than another one if their difference is p_j or more.

- v_j denotes the veto threshold for criterion $g_j(.)$: if an alternative is much worse than the reference profile (by a difference greater than v_j) according to $g_j(.)$, this criterion vetoes the conclusion that the alternative is at least as good as the profile (even if it is very good in other criteria, this difference cannot be compensated).

- k_j denotes the weight of criterion $g_j(.)$; by convention, weights are non-negative numbers and their sum is equal to 1.

- λ (cutting level) denotes the required majority level to warrant an outranking conclusion. For instance, setting λ=0.67 requires that the criteria in favor of the outranking represent at least 67% of the total weight of the criteria (in the absence of veto).

Based on these inputs and parameters the classification of an alternative is obtained performing the following computations:

- $c_j(a_i,b^h)$ is a single-criterion concordance index that indicates how much $g_j(.)$ agrees that a_i outranks (i.e. is at least as good as) b^h, on a scale from 0 (does not agree) to 1 (fully agrees):

$$c_j(a_i,b^h) = \begin{cases} 0, & \text{if } g_j(a_i)-g_j(b^h) < -p_j \\ \left(g_j(a_i)-g_j(b^h)+p_j\right)\big/\left(p_j-q_j\right), & \text{if } -p_j \leq g_j(a_i)-g_j(b^h) < -q_j \\ 1, & \text{if } g_j(a_i)-g_j(b^h) \geq -q_j \end{cases}$$

- $d_j(a_i,b^h)$ is a single-criterion discordance index that indicates whether the performance on $g_j(.)$ vetoes the assertion that a_i outranks b^h (0 means No, 1 means Yes):

$$d_j(a_i,b^h) = \begin{cases} 0, & \text{if } g_j(a_i)-g_j(b^h) > -v_j \\ 1, & \text{if } g_j(a_i)-g_j(b^h) \leq -v_j \end{cases}$$

- $c(a_i,b^h)$ is a global concordance index for the assertion that a_i outranks b^h, taking into account the criteria weights:

$$c(a_i,b^h) = \sum_{j=1}^{n} k_j c_j(a_i,b^h)$$

- $s(a_i,b^h)$ is the overall credibility of the assertion that a_i outranks b^h, summarizing the arguments for and against it:

$$s(a_i,b^h) = c(a_i,b^h).\left[1 - \max_{j \in \{1,\dots,n\}} d_j(a_i,b^h)\right]$$

According to the variant used in this work, an alternative is sorted into C^1 if it is not good enough to outrank b^1, it is sorted into C^2 if it is good enough to outrank b^1 but not good enough to outrank b^2, and so on. Thus, the classification rule is the following:

a_i is classified in category C^h if and only if $s(a_i,b^{h-1}) \geq \lambda$ and $s(a_i,b^h) < \lambda$.

Appendix B. Intermediate results considering all respondents (Please see the Supplementary Material file).

Appendix C. Intermediate results considering the Government perspective (Please see the Supplementary Material file)

Appendix D. Intermediate results considering the Business perspective (Please see the Supplementary Material file)

Appendix E. Intermediate results considering the Knowledge perspective (Please see the Supplementary Material file)

Acknowledgements

References

[1]   Keeney RL. Framing public policy decisions. Int J Technol Policy Manag 2004;4:95–115.

[2]   Marinakis V, Doukas H, Xidonas P, Zopounidis C. Multicriteria decision support in local energy planning: An evaluation of alternative scenarios for the Sustainable Energy Action Plan. Omega 2017;69:1–16.

[3]   Schuwirth N, Reichert P, Lienert J. Methodological aspects of multi-criteria decision analysis for policy support : A case study on pharmaceutical removal from hospital wastewater. Eur J Oper Res 2012;220:472–83.

[4]   Stewart TJ, French S, Rios J. Integrating multicriteria decision analysis and scenario planning-Review and extension. Omega 2013;41:679–88.

[5]   Belton V, Stewart T. Multiple Criteria Decision Analysis. Kluwer Academic Publishers; 2001.

[6]   Greco S, Ehrgott M, Figueira JR. Multiple Criteria Decision Analysis - State of the Art Surveys. New York: Springer-Verlag; 2016.

[7]   Ishizaka A, Nemery P. Multi-criteria Decision Analysis: Methods and Software. Chichester: Wiley; 2013.

[8]   Bana e Costa CA, Lourenço JC, Oliveira MD, Bana e Costa JC. A Socio-technical Approach for Group Decision Support in Public Strategic Planning: The Pernambuco PPA Case. Gr Decis Negot 2014;23:5–29.

[9]   Mastorakis K, Siskos E. Value focused pharmaceutical strategy determination with multicriteria decision analysis techniques. Omega 2016;59:84–96.

[10]   Figueira J, Greco S, Roy B, Slowinski R. An Overview of ELECTRE Methods and their Recent Extensions. J Multi-Criteria Decis Anal 2013;20:61–85.

[11]   Govindan K, Jepsen MB. ELECTRE: A comprehensive literature review on

methodologies and applications. Eur J Oper Res 2016;250:1–29.

[12]   Kumar V, Del Vasto-Terrientes L, Valls A, Schuhmacher M. Adaptation strategies for water supply management in a drought prone Mediterranean river basin: Application of outranking method. Sci Total Environ 2015;540:344–57.

[13]   Neves LP, Martins AG, Antunes CH, Dias LC. A multi-criteria decision approach to sorting actions for promoting energy efficiency. Energy Policy 2008;36:2351–63.

[14]   Van Amsterdam JGC, Best W, Opperhuizen A, De Wolff FA. Evaluation of a procedure to assess the adverse effects of illicit drugs. Regul Toxicol Pharmacol 2004;39:1–4.

[15]   Bui TX. Co-oP - A Group Decision Support System for Cooperative Multiple Criteria Group Decision Making. Berlin Heidelberg: Springer-Verlag; 1987.

[16]   Rolland A. Reference-based preferences aggregation procedures in multi-criteria decision making. Eur J Oper Res 2013;225:479–86.

[17]   Checkland P, Scholes J. Soft Systems Methodology in Action. Chichester: Wiley; 1990.

[18]   Mingers J, Rosenhead J. Problem structuring methods in action. Eur J Oper Res 2004;152:530–54.

[19]   von Winterfeldt D, Fasolo B. Structuring decision problems: A case study and reflections for practitioners. Eur J Oper Res 2009;199:857–66.

[20]   Roy B. Multicriteria Methodology for Decision Aiding. Dordrecht: Kluwer Academic; 1996.

[21]   Neves LP, Dias LC, Antunes CH, Martins AG. Structuring an MCDA model using SSM: A case study in energy efficiency. Eur J Oper Res 2009;199:834–45.

[22]   Keeney RL. Value-Focused Thinking. Cambridge, Massachusetts: Harvard University Press; 1992.

[23]   Antunes CH, Dias L, Dantas G, Mathias J, Zamboni L. An application of Soft Systems Methodology in the evaluation of policies and incentive actions to promote technological innovations in the electricity sector. Energy Procedia 2016;(in print).

[24]   De Smet Y, Nemery P, Selvaraj R. An exact algorithm for the multicriteria ordered clustering problem. Omega 2012;40:861–9.

[25]   Yu W. ELECTRE TRI: Aspects méthodologiques et manuel d'utilisation. Document du LAMSADE n° 74. Paris: Université de Paris–Dauphine. 1992.

[26]   Covas MT, Silva CA, Dias LC. Multicriteria decision analysis for sustainable data centers location. Int Trans Oper Res 2013;20:269–99.

[27]   Sánchez-Lozano JM, Henggeler Antunes C, García-Cascales MS, Dias LC. GIS-based photovoltaic solar farms site selection using ELECTRE-TRI: Evaluating the case for Torre Pacheco, Murcia, Southeast of Spain. Renew Energy 2014;66:478–94.

[28]   Almeida-Dias J, Figueira JR, Roy B. ELECTRE TRI-C : A multiple criteria sorting method based on characteristic reference actions. Eur J Oper Res 2010;204:565–80.

[29]   Almeida-Dias J, Figueira JR, Roy B. A multiple criteria sorting method where each category is characterized by several reference actions: The Electre Tri-nC method. Eur J Oper Res 2012;217:567–79.

[30]   Corrente S, Greco S, Słowiński R. Multiple Criteria Hierarchy Process for ELECTRE Tri methods. Eur J Oper Res 2016;252:191–203.

[31]   Kadziński M, Tervonen T, Rui Figueira J. Robust multi-criteria sorting with the outranking preference model and characteristic profiles. Omega 2015;55:126–40.

[32]   Roy B, Figueira JR, Almeida-Dias J. Discriminating thresholds as a tool to cope with imperfect knowledge in multiple criteria decision aiding: Theoretical results and practical issues. Omega 2014;43:9–20.

[33]   Dalkey N, Helmer O. An Experimental Application of the Delphi Method to the Use of Experts. Manage Sci 1963;9:458–67.

[34] Öztürk M, Tsoukiàs A, Guerrand S. On the Use of a Multicriteria Decision Aiding Tool for the Evaluation of Comfort. In: Bisdorff R, Dias LC, Meyer P, Mousseau V, Pirlot M, editors. Eval. Decis. Model. with Mult. Criteria Case Stud., Berlin, Heidelberg: Springer Berlin Heidelberg; 2015, p. 475–500.

[35] Kirkwood CW, Sarin RK. Ranking with Partial Information: A Method and an Application. Oper Res 1985;33:38–48.

[36] Dias LC. A note on the role of robustness analysis in decision-aiding processes. In: Roy B, Ali Aloulou M, Kalaï R, editors. Robustness OR-DA, Ann. du LAMSADE, No. 7, Paris: Université-Paris Dauphine; 2007, p. 53–70.

[37] Roy B. A missing link in OR-DA: robustness analysis. Found Comput Decis Sci 1998;23:141–60.

[38] Dias LC, Clímaco JN. ELECTRE TRI for groups with imprecise information on parameter values. Gr Decis Negot 2000;9:355–77.

[39] Kadziński M, Tervonen T. Robust multi-criteria ranking with additive value models and holistic pair-wise preference statements. Eur J Oper Res 2013;228:169–80.

[40] Tervonen T, Lahdelma R, Almeida Dias J, Figueira J, Salminen P. SMAA-TRI. In: Linkov I, Kiker GA, Wenning RJ, editors. Environ. Secur. Harb. Coast. Areas Manag. Using Comp. Risk Assess. Multi-Criteria Decis. Anal., Dordrecht: Springer Netherlands; 2007, p. 217–31.

[41] Tervonen T. JSMAA: open source software for SMAA computations. Int J Syst Sci 2014;45:69–81.

[42] Bouyssou D, Marchant T. Multiattribute preference models with reference points. Eur J Oper Res 2013;229:470–81.

[43] Dias L, Mousseau V, Figueira J, Clímaco J. An aggregation/disaggregation approach to obtain robust conclusions with ELECTRE TRI. Eur J Oper Res 2002;138:332–48.

[44] Dias LC, Clímaco JN. ELECTRE TRI for Groups with Imprecise Information on Parameter Values. Gr Decis Negot 2000;9.

[45] Damart S, Dias LC, Mousseau V. Supporting groups in sorting decisions: Methodology and use of a multi-criteria aggregation/disaggregation DSS. Decis Support Syst 2007;43.

[46] Govindan K, Kadziński M, Sivakumar R. Application of a novel PROMETHEE-based method for construction of a group compromise ranking to prioritization of green suppliers in food supply chain. Omega (forthcoming). DOI: 10.1016/j.omega.2016.10.004

[47] Mousseau V, Dias L. Valued outranking relations in ELECTRE providing manageable disaggregation procedures. Eur J Oper Res 2004;156.

[48] Gouveia MC, Dias LC, Antunes CH, Boucinha J, Inácio CF. Benchmarking of maintenance and outage repair in an electricity distribution company using the value-based DEA method. Omega 2015;53:104–14.

[49] Tervonen T, van Valkenhoef G, Baştürk N, Postmus D. Hit-And-Run enables efficient weight generation for simulation-based multiple criteria decision analysis. Eur J Oper Res 2013;224:552–9.

[50] Corrente S, Greco S, Slowinski R. Multiple Criteria Hierarchy Process with ELECTRE and PROMETHEE. Omega 2013;41:820–46.

[51] Corrente S, Figueira JR, Greco S, Słowiński R. A robust ranking method extending ELECTRE III to hierarchy of interacting criteria, imprecise weights and stochastic analysis. Omega (forthcoming). DOI: 10.1016/j.omega.2016.11.008

[52] Mousseau V, Dias L. Valued outranking relations in ELECTRE providing manageable disaggregation procedures. Eur J Oper Res 2004;156:467–82.

[53] Dias L, Mousseau V. IRIS: a DSS for multiple criteria sorting problems. J Multi-Criteria Decis Anal 2003;12:285–98.