

1 This is a PDF file of an unedited manuscript that has been accepted for publication in *Omega*. The
2 manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is
3 published in its final form. Please note that during the production process errors may be discovered
4 which could affect the content, and all legal disclaimers that apply to the journal pertain.

5 The final version will be available at: <https://doi.org/10.1016/j.omega.2017.04.004>

6
7
8 **A multi-criteria approach to sort and rank policies based on Delphi**
9 **qualitative assessments and ELECTRE TRI: the case of smart grids in**
10 **Brazil**

11
12 Luis C. Dias ^{a,b}, Carlos Henggeler Antunes ^{a,c}, Guilherme Dantas ^d, Nivalde de Castro ^d,
13 Lucca Zamboni ^e

14 ^a *INESC Coimbra, Coimbra, Portugal*

15 ^b *CeBER and Faculty of Economics, University of Coimbra, Portugal; lmcdias@fe.uc.pt*

16 ^c *Dept. of Electrical and Computer Engineering, University of Coimbra, Portugal;*
17 *ch@deec.uc.pt*

18 ^d *GESEL, Federal University of Rio de Janeiro, Brazil; guilhermecrvg@yahoo.com.br,*
19 *nivalde.castro@gmail.com*

20 ^e *EDP Brasil, Brazil, lucca.zamboni@edpbr.com.br*

21
22 *Abstract:* This work presents a methodology to perform a multi-criteria decision analysis based
23 on qualitative assessments from multiple experts or stakeholders. The application that motivated
24 these methodological developments is also described, consisting of an assessment of policies to
25 foster technological innovations in the electricity sector, namely in a context of evolution to
26 smart grids. Qualitative assessments were elicited from 28 experts using a Delphi process. The
27 assessments concern the impact of policies on different objectives, but also opinions concerning
28 the importance of these objectives. These opinions are translated into constraints on the weights
29 of the corresponding criteria, in a partial information setting. Then, policies are sorted into
30 categories based on a stochastic multi-criteria analysis for the ELECTRE TRI method. A new
31 ranking approach is also developed to be compatible with the sorting results, based on a
32 robustness analysis of ELECTRE TRI credibility indices. Results are discussed considering the
33 entire set of experts and considering subsets of experts representing different perspectives.

34
35 *Keywords:* Multicriteria, Energy, Policy analysis, Delphi, Partial information, Outranking,
36 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.

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.

1 The application-driven methodological developments presented in this paper aim to address two
2 main challenges faced by the project team. One challenge was to obtain an ELECTRE TRI
3 sorting of the policies based on Delphi assessments that used qualitative scales, not only to
4 assess the impact of the policies on different criteria, but also to inquire about the importance of
5 the criteria. According to a search on the Scopus® database only one relevant study that used an
6 ELECTRE method with Delphi was found [14] (“electre” and “delphi” were the search
7 keywords). The Co-oP decision support system offers several methods as independent
8 components that include building ELECTRE outranking graphs, and Delphi and Nominal
9 Group Technique for group discussions [15]. However, none of these approaches use
10 ELECTRE TRI or propose a way to set the criteria weights based on Delphi.

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

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

25

26 **2. Problem structuring**

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

- 31 • What are the objectives of promoting technological innovation in the electricity sector, i.e.
32 the objectives that justify the policies? In this work, each objective corresponds to an
33 evaluation criterion that assesses the impact of each policy on the pursuit of that objective.
- 34 • Which policies are to be evaluated? Each policy corresponds to a potential action to be
35 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].

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

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

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

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

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

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

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

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

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

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

4 Another question to be addressed at the structuring phase was the definition of the type of
5 assessment result sought. The result should not be the selection of the best policy, because there
6 was no reason why only one policy (or any other predefined number of policies) can be
7 implemented. A ranking would be insufficient, because a ranking gives information about how
8 the alternatives compare to each other, but not on their intrinsic merits: it does not indicate if the
9 best alternative in the ranking is good enough to be implemented.

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

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

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

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

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

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

28 **3. Evaluation methodology and process**

29 **3.1. MCDA aggregation model**

30 The MCDA method used in this work needed to meet several requisites. It should be a sorting
31 method, able to separate the policies into categories and such that the assessment of each policy
32 is independent of other policies (as the model should be also applicable to other policies besides
33 P1-P8). Secondly, it should not be a purely compensatory aggregation model in that a very good
34 performance in one criterion is allowed to cancel out a very poor performance on another
35 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 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	
+2 moderately positive impact	Lower bound for C3 (b^2)
+1 slight positive impact	
0 no impact	Lower bound for C2 (b^1)
-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

1 performance of a policy is below the lower profile of a category, then the criterion fully
2 disagrees (concordance is null), otherwise the criterion fully agrees (concordance is 1).

3 Setting non-null indifference and preference thresholds allows considering a gradual transition
4 from no concordance to full concordance. For this work the analysts proposed to use an
5 indifference threshold $q_j=0.5$ and a preference threshold $p_j=2.5$ for all the criteria $j=1, \dots, 7$,
6 which was discussed and validated with a few key stakeholders. With these thresholds, the
7 concordance index (Appendix A) fully agrees that policy a_i outranks profile b^h when the
8 performance of a_i is at least as good as the performance of b^h , it partially agrees if the
9 performance of a_i is 1 or 2 levels below the performance of b^h (concordance of 0.75 and 0.25,
10 respectively) and does not agree at all if the performance of a_i is 3 levels below the performance
11 of b^h . These values were discussed at the same time that category profiles were defined, so that
12 implications would be considered acceptable. For instance, to access category C3 “*Implement*
13 *with priority*” (outranking b^2), a slight positive impact yields a modest concordance of 0.25, a
14 moderately positive impact yields a large concordance of 0.75, and a strong positive impact
15 yields a maximum concordance of 1.

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

19 20 **3.2. Delphi elicitation process**

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

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

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

1 Concerning the impact of the policies on the objectives, respondents answered the question:
2 “Taking into account the objective of [*Name of the objective*] indicate what is the impact that
3 you attribute to each one of the following policies”. The answers were provided using the scale -
4 5 to +5 presented in Table 1. Since this scale is an ordinal qualitative scale, the median, rather
5 than an average, was used to summarize the answers of the respondents. This median was
6 slightly adjusted by removing 0.5 for a few cases in which the number and levels of answers
7 strictly below the median was much more relevant than the number and levels of answers
8 strictly above the median. For instance, the evaluation of policy P5 in Objective 1 had a median
9 level +3 (chosen by 26% of the respondents), but 48% attributed a lower level (including a
10 negative level for 4%), while 26% attributed a higher level (22% level +4, 4% level +5). In this
11 case, the level most appropriate to represent the group’s opinion was considered to be +2.5. This
12 type of adjustment was made for 8 of the 56 evaluations.

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

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

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

36

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:

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

where ω_l ($l = 0, \dots, 5$) 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 $\omega_0 \leq \omega_1 \leq \dots \leq \omega_5$ 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 $l = 0, \dots, 5$. 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 $l = 0, \dots, 5$, 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

1 weight is 0.49). It would have been possible to simply set the constraints $k_j \leq 0.49$, but the
2 strategy followed not only ensures this inequality, it also ensures no criterion has null weight,
3 which was considered a desirable feature.

4 Bounds for the cutting level λ (Appendix A) were discussed at an informal academic meeting
5 where some stakeholders were present to analyze preliminary results. Demanding unanimity,
6 (i.e. all seven criteria in favor, with $\lambda=7/7$) was deemed unnecessarily stringent. The discussion
7 led to constrain the cutting level to lie between $4/7$ and $5/7$. The lower bound corresponds to the
8 support of a coalition of criteria that represents over 57.1% of the criteria weights. For a vector
9 of equal weights, this requires the support of 4 out of the seven criteria, although the criteria
10 weight bounds allow this level to be achieved by two single criteria. The upper bound
11 corresponds to the support of a coalition of criteria that represents over 71.4% of the criteria
12 weights, which was deemed to be a comfortable majority. Since these limits are different from
13 the usually “round” limits set by decision makers (e.g., $1/2$, $2/3$, $3/4$), results are also presented
14 considering an interval $\lambda \in [1/2, 3/4]$.

15 In the preliminary presentation of the sorting results, it was considered that stakeholders would
16 also welcome a ranking of the policies. This raised an interesting question, since a ranking
17 procedure such as ELECTRE III [10] would not be appropriate. Indeed, such procedures would
18 compare the policies against each other to obtain a partial ranking that might contradict the
19 sorting results. Moreover, adding or deleting a policy might change the relative position of the
20 remaining policies. The solution would need to be a reference-based ranking approach, based on
21 sorting profiles, a type of approach that has been proposed and studied only in recent years
22 [16,42].

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

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

35 Let $W(a_i)$ and $B(a_i)$, respectively, denote the indices of the worst and best category where policy
36 a_i can be sorted when satisfying the constraints placed on the parameter values ($i=1, \dots, m$). If the

1 interval of variation for the cutting level is $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, then $W(a_i)$ is the highest value
 2 $h \in \{1, \dots, k\}$ such that $s_m(a_i, b^{h-1}) \geq \lambda_{\max}$. On the other hand, $B(a_i)$ is the lowest value $h \in \{1, \dots, k\}$
 3 such that $s^M(a_i, b^h) < \lambda_{\min}$. Let us note that in some circumstances there may be categories in-
 4 between $W(a_i)$ and $B(a_i)$ where a_i cannot be sorted [43].

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

7 Pr1) Policies are ranked according to the worst category they can have: if $W(a_i) > W(a_j)$, then a_j
 8 should have a worse rank. As a particular case, policies robustly sorted into C^h (i.e., with
 9 $W(a_i) = B(a_i) = h$) are better ranked than those robustly sorted into a worse category C^q ($q < h$).

10 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
 11 reach a better category than a_j , i.e. $B(a_i) > B(a_j)$, then a_j should have a worse rank.

12 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
 13 the same best category $B(a_i) = B(a_j) = \beta$, a lexicographic comparison of vectors $(s_m(a_i, b^{\omega-1}),$
 14 $\dots, s_m(a_i, b^\beta))$ and $(s_m(a_j, b^{\omega-1}), \dots, s_m(a_j, b^\beta))$ is used to obtain a more discriminating ranking,
 15 with the rationale that the better ranked alternative would be last one to worsen its worst
 16 category if λ_{\max} increases:

$$17 \quad a_i \text{ is ranked better than } a_j \Leftrightarrow [s_m(a_i, b^{\omega-1}) > s_m(a_j, b^{\omega-1})] \vee$$

$$18 \quad \exists g \in [\omega, \beta]: s_m(a_i, b^g) = s_m(a_j, b^g) \text{ for all } h = \omega-1, \dots, g-1 \wedge s_m(a_i, b^g) > s_m(a_j, b^g)].$$

20 It is debatable whether a policy with $W(a_i) = B(a_i) = 3$ should be better ranked than a policy with a
 21 $W(a_j) = 2$ and $B(a_j) = 4$. This rule focuses on the worst case, which adheres to the “pessimistic” or
 22 “conservative” character of the ELECTRE TRI variant used in this work. It would be equally
 23 possible to follow an optimistic or “benefit of doubt” procedure as follows:

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

26 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
 27 in a worse category than a_j , i.e. $W(a_i) < W(a_j)$, then a_j should have a better rank.

28 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
 29 the same best category $B(a_i) = B(a_j) = \beta$, a lexicographic comparison of vectors
 30 $(s^M(a_i, b^\beta), \dots, s^M(a_i, b^{\omega-1}))$ and $(s^M(a_j, b^\beta), \dots, s^M(a_j, b^{\omega-1}))$ is used, since the best category is
 31 defined by a maximum credibility.

32 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
 33 this inequalities is strict, then a_i is better ranked than a_j .

34

1 **4. Results**

2 **4.1. Results considering all respondents**

3 This subsection reports the results considering the answers of all 28 respondents who
 4 participated in the two rounds of the Delphi process described in Section 3. Concerning the
 5 impact of the policies on the different objectives the results are presented in Table 2. Referring
 6 to the descriptions in Table 1, respondents consider that all policies have a positive impact for
 7 the different objectives, and it is clear that no policy will be sorted into category C^l
 8 (*Uninteresting*). This corresponds to the expectation of the project team had when conceiving
 9 the policies to be evaluated.

10
 11 *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

12
 13 Since the profiles defining the categories and the indifference and preference thresholds have
 14 been fixed (Section 3.1), it is now possible to compute single-criterion concordance indices, as
 15 described in Appendix A. These indices indicate how much each criterion agrees with an
 16 outranking relation, from a minimum of 0 to a maximum of 1. The evaluations provided by the
 17 respondents are sufficient, with two exceptions (Policy P1 on objective O1 and policy P5 on
 18 objective O5), to guarantee a strong support for a policy to be at least sorted into category C^3 ,
 19 with concordance 0.75 or higher (Table B1 in Appendix B). Concordance indices for attaining
 20 category C^4 are obviously lower, but still strong in many cases (Table B2 in Appendix B).

21 Concerning discordance, respondents were asked to indicate what impact levels they would
 22 consider negative enough to discard implementation of a policy. According to the median
 23 responses in the second round of the Delphi process, this level would be -1.5 for the first
 24 criterion (*Environ. & human health*) and it would be -2 for the remaining criteria. Accordingly,
 25 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 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,

1 O6 and O7 were strongly important (level 4), and O2 was extremely important (level 5). This
 2 originated the following constraints:

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

4 The value of ε was arbitrarily set to 0.01 to enforce a strict inequality. A mathematical program
 5 approach was then used to maximize and minimize $s(a_i, b^h)$ for different category profiles (for
 6 details, see [47]). The results are presented in Table 4. Except P4, all policies could be sorted
 7 into C3 with unanimity (P4 is also acceptable in that category for 61% of the respondents), and
 8 it is also unanimous that P2, P4 and P6 could be sorted into C4 (P7 is also acceptable in that
 9 category for 71% of the respondents). These results yield a clear picture of what classifications
 10 are not supported by the group, but do not offer much discrimination between the policies
 11 (Damart et al. [45] proposed an iterative process of multiple discussion rounds to obtain a more
 12 clear-cut result).

13

14 *Table 4. Proportion of respondents that would accept each potential classification considering*
 15 *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%	

16

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

24 A strict consensus about the ranking (or a partial ranking) of the weights did not arise from the
 25 responses, i.e., there are no two criteria g_j and g_k such that all the respondents assigned a level of
 26 importance to g_j which was higher than the level they assigned to g_k . However, if one compares,

1 for instance, criteria g_4 and g_5 it is possible to notice the following: 100% of the respondents
 2 agree that g_4 is at least level 2 (and the same applies to g_5); 100% of the respondents agree that
 3 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
 4 least level 4 (but only 61% of them for g_5); and finally, 29% of the respondents agree that g_4 is
 5 level 5 (vs. 18% for g_5). As explained in section 3.3, this provides a compelling rationale to
 6 require $k_4 > k_5$. Similar conclusions can be made concerning other pairs of criteria, based on the
 7 cumulative percentage of responses supporting that the criteria importance level is at least l (for
 8 $l = 0, \dots, 5$) (Figure 1). Figure 2a) depicts the resulting partial ranking of the criteria weights.

9 A few additional weaker inequalities were added, to reflect other pairs for which an examination
 10 of the individual responses (not presented here) indicated there was a large agreement: $k_6 \geq k_2$
 11 (82% of the respondents agreed) and $k_3 \geq k_5$ (89% agreed), implying by transitivity $k_6 \geq k_4$ (96%
 12 agreed) and $k_6 \geq k_5$ (93% agreed). The resulting relation is depicted in Figure 2b). Contrarily to
 13 the relations in Figure 2a), which were considered as strict inequalities (a minimum arbitrarily
 14 small difference of 0.005 was required), $k_6 \geq k_2$ and $k_3 \geq k_5$ were considered as weak inequalities
 15 (admitting equality). Finally, to prevent a criterion from having a null weight, a minimum
 16 weight was required. Since in this case the minimum weight cannot be set based on limits to
 17 trade-offs (as in, e.g. [48]), the lowest minimum considered reasonable was chosen to be 0.085,
 18 so that no criterion would have null weight or a weight higher than $1 - 6 \times 0.085 = 0.49$.

19

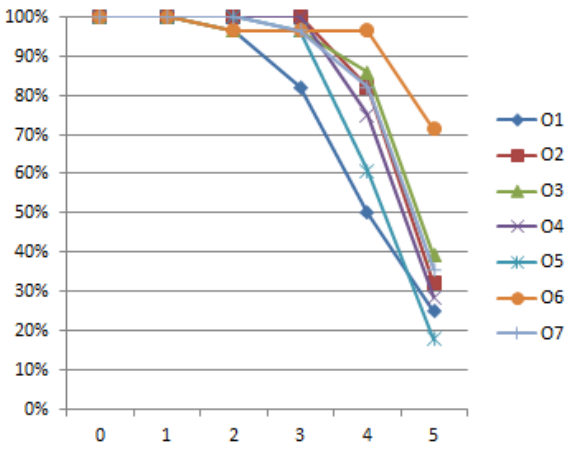


Figure 1. Cumulative percentage of responses supporting that the criteria importance level is at least l (for $l = 0, \dots, 5$).

20

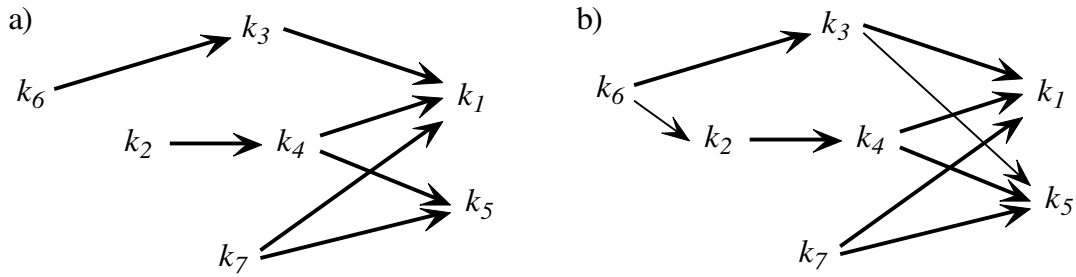


Figure 2. Partial ranking of the criteria weights (all respondents): a) based on cumulative responses; b) based on cumulative responses and weaker relations.

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)

1 *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

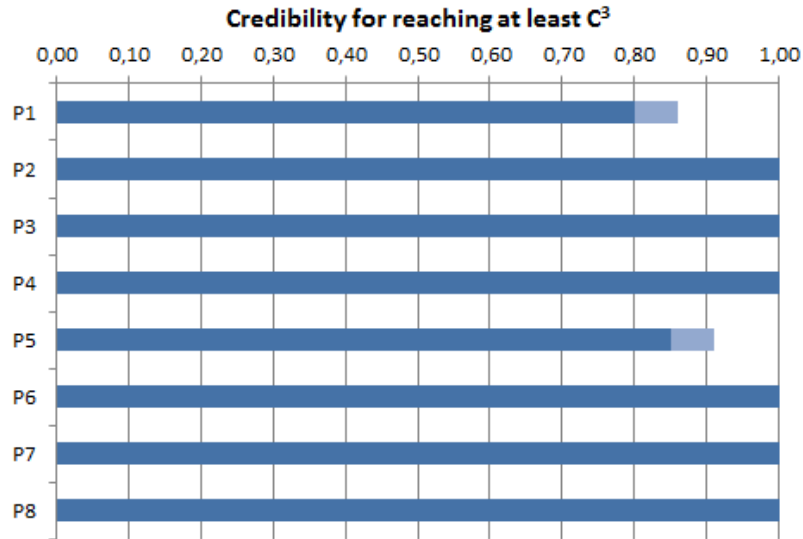
2 ^(*)Sorting varies between “Implement with priority” and “Implement with maximum priority”

3

4 Table 6 presents the sorting results in a way that facilitates communication with stakeholders
 5 making use of a label *Implement with high priority* to represent the interval $[C^3, C^4]$, i.e.
 6 hesitation between C^3 and C^4 . Indeed, policies in the interval $[C^3, C^4]$ can be considered as
 7 having more priority than policies in C^3 and having less priority than policies in C^4 . Results
 8 were also obtained considering a wider variation interval of $[1/2, 3/4]$ for the cutting level, to
 9 assess the extent to which conclusions would change (Table 5, values in parenthesis). By
 10 accepting a more demanding value for the cutting level, 0.75, lower classifications become
 11 possible, and this occurs only for P4 in 9.9% of the cases. On the other hand, by accepting a less
 12 demanding value for the cutting level, 0.50, higher classifications become possible, and this
 13 occurs only for P3 and P8 in 8.4% of the cases. The remaining minimum and maximum
 14 classifications would not change.

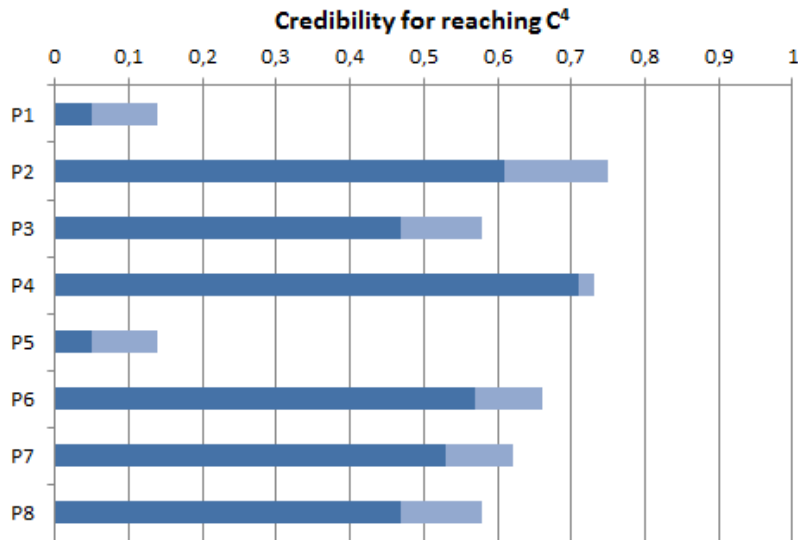
15 Finally, a ranking of the policies within each category or interval of categories was obtained as
 16 described in Section 3.4. Figure 3 depicts the credibility that a policy may reach category C^3
 17 (i.e. credibility of outranking profile b^2), whereas Figure 4 depicts the credibility that a policy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

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



1

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



4

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

7

8 *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	<i>(no need for a tie break)</i>	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

9

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.

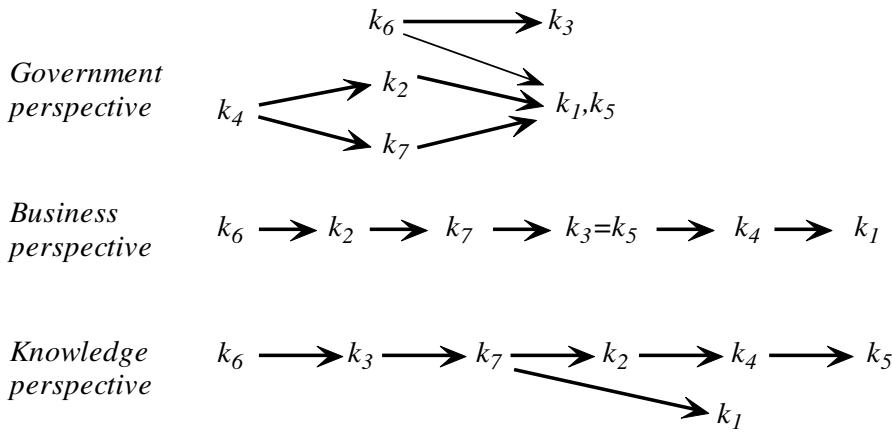
1 Tables C5, D5 and E5 in Appendices C, D and E summarize the Delphi responses concerning
 2 the importance of each objective, from which the partial rankings of criteria weights were
 3 derived. The only contradictory opinion found in these relations (Figure 5) concerns the
 4 importance of the second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

6

7 *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

8



9

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

13

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

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

24

1 *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

2

3 *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

4 (*) *ex-aequo*

5

6 Overall, the sorting for each perspective is not very different from the sorting obtained
7 considering the entire set of responses (Table 6). According to all the perspectives analyzed, all
8 policies have merits for implementation with some degree of priority. The only exception is
9 policy P5 under the *Business perspective*, where it could belong to category C² (Wait and see)
10 for some of the parameter values accepted. This policy, together with P1, is at the bottom of the
11 ranking for the three perspectives. All perspectives rank policy P4 at the top. Comparing the
12 different perspectives, results are in agreement to a large extent since there is always a category
13 in common for every policy, except P6.

14

15 **5. Discussion and conclusions**

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

20 The methodology followed in this work is based on a well-known and proven multi-criteria
21 sorting method: ELECTRE TRI (pseudo-conjunctive variant). Expert judgment from 28

1 stakeholders was obtained by means of a Delphi survey eliciting inputs on a qualitative scale
2 about policy impacts and some of the main parameters of the ELECTRE TRI model: criteria
3 weights and veto thresholds. However, the methodology needed to be developed to meet the
4 specific requirements of this application.

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

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

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

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

33 A limitation of this work is that it does not address potential synergies among policies, as
34 policies were evaluated independently of each other considering their intrinsic advantages and
35 disadvantages. If it was necessary to consider synergies between a few policies only, the same
36 methodology could be used to evaluate composite policies (e.g. a policy P9 could represent

1 implementing two other policies at the same time), but evaluating all combinations might be
2 impractical. Future research might however assess synergies between the most promising
3 policies (according to the outputs of this study). An interesting extension of this methodology
4 would be to view the problem as a portfolio evaluation one, where each portfolio would consist
5 of multiple policies.

6 The research conducted in this work may also be repeated in different contexts, or even in the
7 same context but with different stakeholders. Indeed, the results are contingent on the views of
8 the experts who collaborated in this work. The larger the number and the variety of stakeholders
9 involved, the more representative results will be. In one of the public presentations of this work
10 it was suggested that a fourth perspective could be that of the general public (who are mainly
11 consumers). Consumer representatives might not necessarily be experts capable of assessing the
12 impacts of policies on the different objectives, but they could express their views on the
13 questions concerning the importance of the criteria and the possibility of veto. Future
14 applications of this methodology can then use different sets of respondents to answer different
15 questions in the Delphi survey, thus allowing the elicitation of the views of non-experts
16 concerning aspects that do not require domain-specific expertise.

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

21 22 **Appendix A. ELECTRE TRI computations**

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

- 25 • $A=\{a_1,\dots,a_m\}$ is the set of m alternatives to be sorted.
- 26 • $G=\{g_1(\cdot),\dots,g_n(\cdot)\}$ is a set of n evaluation criteria. The following computations assume that
27 (as is the case in this work), the higher is $g_j(a_i)$, the better is the performance of alternative a_i
28 on criterion $g_j(\cdot)$.
- 29 • $B=\{b^0,\dots,b^k\}$ denotes the set of profiles that define a set of k categories, $\{C^1,\dots,C^k\}$. Each
30 category C^h ($h=1,\dots,k$) is defined through two reference profiles: a lower bound b^{h-1} and an
31 upper bound b^h . By convention, C^1 denotes the worst category and C^k denotes the best
32 category.

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

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

- 1 • p_j denotes the preference threshold for criterion $g_j(\cdot)$: one level of performance is considered
2 to be undoubtedly better than another one if their difference is p_j or more.
- 3 • v_j denotes the veto threshold for criterion $g_j(\cdot)$: if an alternative is much worse than the
4 reference profile (by a difference greater than v_j) according to $g_j(\cdot)$, this criterion vetoes the
5 conclusion that the alternative is at least as good as the profile (even if it is very good in
6 other criteria, this difference cannot be compensated).
- 7 • k_j denotes the weight of criterion $g_j(\cdot)$; by convention, weights are non-negative numbers
8 and their sum is equal to 1.
- 9 • λ (cutting level) denotes the required majority level to warrant an outranking conclusion.
10 For instance, setting $\lambda=0.67$ requires that the criteria in favor of the outranking represent at
11 least 67% of the total weight of the criteria (in the absence of veto).

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

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

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

- 17 • $d_j(a_i, b^h)$ is a single-criterion discordance index that indicates whether the performance on
18 $g_j(\cdot)$ vetoes the assertion that a_i outranks b^h (0 means No, 1 means Yes):

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

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

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

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

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

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

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

5
6 **Appendix B. Intermediate results considering all respondents** (Please see the Supplementary
7 Material file).

8
9 **Appendix C. Intermediate results considering the Government perspective** (Please see the
10 Supplementary Material file)

11
12 **Appendix D. Intermediate results considering the Business perspective** (Please see the
13 Supplementary Material file)

14
15 **Appendix E. Intermediate results considering the Knowledge perspective** (Please see the
16 Supplementary Material file)

17 **Acknowledgements**

18
19 The authors acknowledge the anonymous reviewers for their valuable comments on a previous
20 version of this article. This work has been supported by EDP (Bandeirante Energia and Espírito
21 Santo Centrais Elétricas) project “Evaluation of policies and incentive actions for technological
22 innovations in the electricity sector: analysis of the international experience and proposals for
23 Brazil” part of the R&D Program regulated by ANEEL, Brazil. The first two authors also
24 acknowledge the support of the Portuguese Foundation for Science and Technology under
25 project UID/MULTI/00308/2013 and POCI-01-0145-FEDER-016765. The authors are grateful
26 to the 28 experts who participated in the Delphi survey.

27 **References**

- 28
29 [1] Keeney RL. Framing public policy decisions. *Int J Technol Policy Manag* 2004;4:95–
30 115.
31 [2] Marinakis V, Doukas H, Xidonas P, Zopounidis C. Multicriteria decision support in
32 local energy planning: An evaluation of alternative scenarios for the Sustainable Energy
33 Action Plan. *Omega* 2017;69:1–16.
34 [3] Schuwirth N, Reichert P, Lienert J. Methodological aspects of multi-criteria decision
35 analysis for policy support : A case study on pharmaceutical removal from hospital
36 wastewater. *Eur J Oper Res* 2012;220:472–83.
37 [4] Stewart TJ, French S, Rios J. Integrating multicriteria decision analysis and scenario

- 1 planning-Review and extension. *Omega* 2013;41:679–88.
- 2 [5] Belton V, Stewart T. *Multiple Criteria Decision Analysis*. Kluwer Academic Publishers;
3 2001.
- 4 [6] Greco S, Ehrgott M, Figueira JR. *Multiple Criteria Decision Analysis - State of the Art*
5 *Surveys*. New York: Springer-Verlag; 2016.
- 6 [7] Ishizaka A, Nemery P. *Multi-criteria Decision Analysis: Methods and Software*.
7 Chichester: Wiley; 2013.
- 8 [8] Bana e Costa CA, Lourenço JC, Oliveira MD, Bana e Costa JC. A Socio-technical
9 Approach for Group Decision Support in Public Strategic Planning: The Pernambuco
10 PPA Case. *Gr Decis Negot* 2014;23:5–29.
- 11 [9] Mastorakis K, Siskos E. Value focused pharmaceutical strategy determination with
12 multicriteria decision analysis techniques. *Omega* 2016;59:84–96.
- 13 [10] Figueira J, Greco S, Roy B, Slowinski R. An Overview of ELECTRE Methods and their
14 Recent Extensions. *J Multi-Criteria Decis Anal* 2013;20:61–85.
- 15 [11] Govindan K, Jepsen MB. ELECTRE: A comprehensive literature review on
16 methodologies and applications. *Eur J Oper Res* 2016;250:1–29.
- 17 [12] Kumar V, Del Vasto-Terrientes L, Valls A, Schuhmacher M. Adaptation strategies for
18 water supply management in a drought prone Mediterranean river basin: Application of
19 outranking method. *Sci Total Environ* 2015;540:344–57.
- 20 [13] Neves LP, Martins AG, Antunes CH, Dias LC. A multi-criteria decision approach to
21 sorting actions for promoting energy efficiency. *Energy Policy* 2008;36:2351–63.
- 22 [14] Van Amsterdam JGC, Best W, Opperhuizen A, De Wolff FA. Evaluation of a procedure
23 to assess the adverse effects of illicit drugs. *Regul Toxicol Pharmacol* 2004;39:1–4.
- 24 [15] Bui TX. *Co-oP - A Group Decision Support System for Cooperative Multiple Criteria*
25 *Group Decision Making*. Berlin Heidelberg: Springer-Verlag; 1987.
- 26 [16] Rolland A. Reference-based preferences aggregation procedures in multi-criteria
27 decision making. *Eur J Oper Res* 2013;225:479–86.
- 28 [17] Checkland P, Scholes J. *Soft Systems Methodology in Action*. Chichester: Wiley; 1990.
- 29 [18] Mingers J, Rosenhead J. *Problem structuring methods in action*. *Eur J Oper Res*
30 2004;152:530–54.
- 31 [19] von Winterfeldt D, Fasolo B. Structuring decision problems: A case study and
32 reflections for practitioners. *Eur J Oper Res* 2009;199:857–66.
- 33 [20] Roy B. *Multicriteria Methodology for Decision Aiding*. Dordrecht: Kluwer Academic;
34 1996.
- 35 [21] Neves LP, Dias LC, Antunes CH, Martins AG. Structuring an MCDA model using
36 SSM: A case study in energy efficiency. *Eur J Oper Res* 2009;199:834–45.
- 37 [22] Keeney RL. *Value-Focused Thinking*. Cambridge, Massachusetts: Harvard University
38 Press; 1992.
- 39 [23] Antunes CH, Dias L, Dantas G, Mathias J, Zamboni L. An application of Soft Systems
40 Methodology in the evaluation of policies and incentive actions to promote technological
41 innovations in the electricity sector. *Energy Procedia* 2016;(in print).
- 42 [24] De Smet Y, Nemery P, Selvaraj R. An exact algorithm for the multicriteria ordered
43 clustering problem. *Omega* 2012;40:861–9.
- 44 [25] Yu W. *ELECTRE TRI: Aspects méthodologiques et manuel d'utilisation*. Document du
45 LAMSADE n° 74. Paris: Université de Paris–Dauphine. 1992.
- 46 [26] Covas MT, Silva CA, Dias LC. Multicriteria decision analysis for sustainable data
47 centers location. *Int Trans Oper Res* 2013;20:269–99.
- 48 [27] Sánchez-Lozano JM, Henggeler Antunes C, García-Cascales MS, Dias LC. GIS-based
49 photovoltaic solar farms site selection using ELECTRE-TRI: Evaluating the case for

- 1 Torre Pacheco, Murcia, Southeast of Spain. *Renew Energy* 2014;66:478–94.
- 2 [28] Almeida-Dias J, Figueira JR, Roy B. ELECTRE TRI-C : A multiple criteria sorting
3 method based on characteristic reference actions. *Eur J Oper Res* 2010;204:565–80.
- 4 [29] Almeida-Dias J, Figueira JR, Roy B. A multiple criteria sorting method where each
5 category is characterized by several reference actions: The Electre Tri-nC method. *Eur J*
6 *Oper Res* 2012;217:567–79.
- 7 [30] Corrente S, Greco S, Słowiński R. Multiple Criteria Hierarchy Process for ELECTRE
8 Tri methods. *Eur J Oper Res* 2016;252:191–203.
- 9 [31] Kadziński M, Tervonen T, Rui Figueira J. Robust multi-criteria sorting with the
10 outranking preference model and characteristic profiles. *Omega* 2015;55:126–40.
- 11 [32] Roy B, Figueira JR, Almeida-Dias J. Discriminating thresholds as a tool to cope with
12 imperfect knowledge in multiple criteria decision aiding: Theoretical results and
13 practical issues. *Omega* 2014;43:9–20.
- 14 [33] Dalkey N, Helmer O. An Experimental Application of the Delphi Method to the Use of
15 Experts. *Manage Sci* 1963;9:458–67.
- 16 [34] Öztürk M, Tsoukiàs A, Guerrand S. On the Use of a Multicriteria Decision Aiding Tool
17 for the Evaluation of Comfort. In: Bisdorff R, Dias LC, Meyer P, Mousseau V, Pirlot M,
18 editors. *Eval. Decis. Model. with Mult. Criteria Case Stud.*, Berlin, Heidelberg: Springer
19 Berlin Heidelberg; 2015, p. 475–500.
- 20 [35] Kirkwood CW, Sarin RK. Ranking with Partial Information: A Method and an
21 Application. *Oper Res* 1985;33:38–48.
- 22 [36] Dias LC. A note on the role of robustness analysis in decision-aiding processes. In: Roy
23 B, Ali Aloulou M, Kalaï R, editors. *Robustness OR-DA*, Ann. du LAMSADE, No. 7,
24 Paris: Université-Paris Dauphine; 2007, p. 53–70.
- 25 [37] Roy B. A missing link in OR-DA: robustness analysis. *Found Comput Decis Sci*
26 1998;23:141–60.
- 27 [38] Dias LC, Clímaco JN. ELECTRE TRI for groups with imprecise information on
28 parameter values. *Gr Decis Negot* 2000;9:355–77.
- 29 [39] Kadziński M, Tervonen T. Robust multi-criteria ranking with additive value models and
30 holistic pair-wise preference statements. *Eur J Oper Res* 2013;228:169–80.
- 31 [40] Tervonen T, Lahdelma R, Almeida Dias J, Figueira J, Salminen P. SMAA-TRI. In:
32 Linkov I, Kiker GA, Wenning RJ, editors. *Environ. Secur. Harb. Coast. Areas Manag.*
33 *Using Comp. Risk Assess. Multi-Criteria Decis. Anal.*, Dordrecht: Springer Netherlands;
34 2007, p. 217–31.
- 35 [41] Tervonen T. JSMAA: open source software for SMAA computations. *Int J Syst Sci*
36 2014;45:69–81.
- 37 [42] Bouyssou D, Marchant T. Multiattribute preference models with reference points. *Eur J*
38 *Oper Res* 2013;229:470–81.
- 39 [43] Dias L, Mousseau V, Figueira J, Clímaco J. An aggregation/disaggregation approach to
40 obtain robust conclusions with ELECTRE TRI. *Eur J Oper Res* 2002;138:332–48.
- 41 [44] Dias LC, Clímaco JN. ELECTRE TRI for Groups with Imprecise Information on
42 Parameter Values. *Gr Decis Negot* 2000;9.
- 43 [45] Damart S, Dias LC, Mousseau V. Supporting groups in sorting decisions: Methodology
44 and use of a multi-criteria aggregation/disaggregation DSS. *Decis Support Syst* 2007;43.
- 45 [46] Govindan K, Kadziński M, Sivakumar R. Application of a novel PROMETHEE-based
46 method for construction of a group compromise ranking to prioritization of green
47 suppliers in food supply chain. *Omega* (forthcoming). DOI:
48 10.1016/j.omega.2016.10.004
- 49 [47] Mousseau V, Dias L. Valued outranking relations in ELECTRE providing manageable
50 disaggregation procedures. *Eur J Oper Res* 2004;156.

- 1 [48] Gouveia MC, Dias LC, Antunes CH, Boucinha J, Inácio CF. Benchmarking of
2 maintenance and outage repair in an electricity distribution company using the value-
3 based DEA method. *Omega* 2015;53:104–14.
- 4 [49] Tervonen T, van Valkenhoef G, Baştürk N, Postmus D. Hit-And-Run enables efficient
5 weight generation for simulation-based multiple criteria decision analysis. *Eur J Oper*
6 *Res* 2013;224:552–9.
- 7 [50] Corrente S, Greco S, Slowinski R. Multiple Criteria Hierarchy Process with ELECTRE
8 and PROMETHEE. *Omega* 2013;41:820–46.
- 9 [51] Corrente S, Figueira JR, Greco S, Słowiński R. A robust ranking method extending
10 ELECTRE III to hierarchy of interacting criteria, imprecise weights and stochastic
11 analysis. *Omega* (forthcoming). DOI: 10.1016/j.omega.2016.11.008
- 12 [52] Mousseau V, Dias L. Valued outranking relations in ELECTRE providing manageable
13 disaggregation procedures. *Eur J Oper Res* 2004;156:467–82.
- 14 [53] Dias L, Mousseau V. IRIS: a DSS for multiple criteria sorting problems. *J Multi-Criteria*
15 *Decis Anal* 2003;12:285–98.