Comparing outcomes profiles of public programmes with ELECTRE assistance instead of observing evolution of a synthetic impact index :illustration of the benefits of a multicriteria evaluation approach

Maurice Baslé

Comparing outcomes profiles of public programmes with ELECTRE assistance instead of observing evolution of a synthetic impact index :illustration of the benefits of a multicriteria evaluation approach

Maurice Baslé

Abstract

This paper will show how we could ameliorate the quality of public policies and programmes monitoring and on going evaluations by using established outranking methods and Electre software assistance..

The craftiness is to take the outcomes of programmes for each period and to elaborate the outcomes multidimensional profile, and then to compare these profiles year after year without having the need to calculate and to observe the evolution of the corresponding synthetic impact index.

This article outlines all of the net advantages of comparing annual outcomes profiles rather than the evolution of theaggregatei index. We aim to show that the outcomes profiles can be compared by the same outranking methods as those used in the case of several options or projects compared during the same year. The well-known Electre software assistance is used.

Key-words: multicriteria evaluation, fuzzy integral, outranking method, impact indexes, public policies and programs outcomes profiles.

Introduction

In all the impact evaluations, assessment of the impact of the public action leads unavoidably to a need of impact synthetic indexes. So, impact indexes are fashion.

But they are very expensive and their Confidence Ratio is generally weak. This article begins with a preliminary question: is the measurement of a public program impact by a synthetic index relevant and fiable? Our questioning is based on **two critiques** of this need to have synthetical impact indicators for an "on going process of Monitoring and Evaluation".

The first critique is that many too synthetic indicators describe an evolution of structural changes and not an evolution of outcomes directly imputable to a public policy. Two good examples of such data are the Gross Domestic Product (GDP) and the Human Development Index (HDI). The evolution of these "descriptors" do not strictly reveal the real outcomes of public policy. They are synthetic indicators that reveal the impact of the evolution of numerous causal factors (for example added value, monetary income, the state of health) that roughly reflect general contexts or structural evolutions, which can themselves be the outcome of many changes, changes in private behaviour for example. The evolution of a too synthetic index contains the partial effects of new public policies but the observed impact is not strictly imputable to one factor, the public program. The evolution of the index can only be initially interpreted at the systemic level because they are the outcomes of private human action as well as public policy effects and structural trend is present. Therefore, we cannot deduce that such proxies are useful to accurately measure the pure effects of the implementations of public policy and their societal and final impact or consequences. This open the door to examine less synthetic effects and more multidimensional data in order to get more strictly imputable effects of a public program.

The second critique applies to the technical construction of this combination of indicators.

The method of aggregation is generally very simple: all the outcomes are given a value based on the same scale, a single criterion of synthesis (that is monetarization of effects, for example Euros). And then, one often uses a simple weighted arithmetic average. There are however, many critics which set themselves against this composition law which is used by many official organisations when aggregating indicators.

The second critique applies to the technical construction of this combination of indicators.

The method is very simply based: all the outcomes are given a value based on the same scale that has a single criterion of synthesis (for example euros, in the case of monetization). Why have these synthetic indexes at all? For an historical index (or when we speak about evolutions), one often uses a simple weighted arithmetic average. There are however, many critics whichset set themselves against this composition law which is used by many official organisations to aggregate indicators (Marchant and al, 2003).

The critics appear to confirm the dissatisfaction caused by over-synthesised data such as the GDP and the HDI (Perret, 2002) , or, in a slightly different way, the dissatisfaction provided by a general ranking (as in the case of the Shanghai University's ranking of Universities around the world). Indeed, the use of a weighted arithmetic average based on just a single synthetic criterion is asking a great deal and the solutions are often weak in responding to the problems met. Moreover this application requires that one respects three principals: independence of the criteria to avoid counting the same evolution twice; normalization of the scales in order to join evolutions; the infinite compensation or substitution between criteria to combine them.

From these observations and difficulties that are well identified in currently available literature (Baslé, 2008), we ask if it is possible to suggest a methodology more appropriate to the monitoring of Ongoing Evaluations of public policy outcomes issues. To show that there are better ways than by the synthetic index of an impact, we suggest coming back to the multicriteria evaluation methods (Bell and al, 1994) and to see how they give an opportunity to compare outcome profiles.

To show that there are better ways to take readings other than by the synthetic index of an impact, we suggest to come back to the multicriteria evaluation methods (Bell and al, 1994)) and to see how they could give an opportunity to compare direct, disaggregated and multridimensional outcomes profiles.

Multicriteria evaluations usually compare, in ex ante evaluations, options or projects relative to a potential or expected net return or outcomes in several dimensions of outcomes. Then, the work of outranking in multicriteria decision aid is to compare each outcome in its dimension, not to produce a comprehensive ranking.

Here, the main intention is to take the outcomes of a policy and its programmes for each period, to create the annual profile and to compare these annual profiles year on year without having to calculate a composite index of synthetic impact. We aim to show that these profiles comparisons can be realized by the same outranking methods as those used in the case of several options or projects compared during the same year.

This article will by the way demonstrate the superiority of comparing profiles in this way, which are therefore less synthetic. It outlines all of the net advantages of comparing less aggregated outcome profiles when the decision-maker’s problem is traditionally to build composite indicators using outranking methods (ELECTRE).

We will show that the comparison of outcome profiles tells us a lot about the outranking of options, and furthermore, we suggest that outcome indices at different periods can be considered as so many options (which are temporal series rendered as a single moment in time). Each period is graded according to different criteria. We refer to well-known outranking methods to compare evolutions of outcome profiles. This is often more than sufficient for the final deliberation, which will be about the progress or the stagnation of the overall trajectory. In this way, the method consists of comparing longer periods and less aggregation is required at the beginning.

We will demonstrate two points. Firstly, the superiority of comparing multicriteria profiles in a single synthetic impact index. Secondly, that this method of comparing profiles can be applied to the analysis of evolutions of outcomes over time. Each annual outcome profile can be considered as an option. The temporal series of the profiles, thanks to the multicriteria evaluation, replaces the evolution of the synthetic impact index. By way of application, we show that it is possible to use decision aid tools such as ELECTRE in an Ongoing Evaluation to build a monitoring composite index of the impact of a public policy. With our methods, an algorithm or a fuzzy integral is not required in the first instance, as comparisons of profiles in time give an overview of the trajectory of outcomes or criteria and an overview of the progress (or the stagnation) of the effects of the public policy.

Note that our proposition with regard to Guitouni and Martel's paper (Guitouni and Martel, 1998) . takes place upstream and downstream of this multicriteria decision aid process. Upstream, it establishes a new methodology that does not require the comparison of ex-ante options but allows for the gathering of multidimensional data for monitoring and Ongoing Evaluation. Downstream, because the method creates the ability to better compare outcome profiles. This opens the door to the design of better recommendations.

1. Why do we generally use synthetic indices to monitor and evaluate public policies and programmes ?

It is essential for good Governance to have relevant public laws to regularly maintain their development and regulation (Kaufmann and al, 2008) . In a knowledge-based society, evaluations and assessments of the effects of public policies have an increasing place (Baslé, 2008). To produce useful metrics to aid in public policy decision making, legislative assemblies and administrators at both State government and local authority level draw up and send out calls for proposals to the scientific and expert community. The purpose of the terms of references of these calls is to formulate the key evaluation questions. Proposals should answer and the deliverables should contain assessment and recommendations.

In order to prove the socio-economic utility of public policies and programmes, the most relevant proposals use and develop intensive use of lessons from “Science and Public Policy” (See for example the Science and Public Policy Review, ISSN 0302-3427, Online ISSN: 1471-5430). The case of Regional authorities in France is representative of the increasing need for proofs for the efficacy of their policies.

One way of helping them to do this is by using cost-benefit analyses which can then be integrated into other works and programmes that have been outlined and based upon specifications laid down by these regional authorities. However, there are many other new methodologies and tools for monitoring and evaluating activities.

Development of evaluation activities induces learning and increases the quality of the evaluation reports (Baslé M., 2008). Now it is more and more important to predict and sometimes to measure, the local impact of public policies and programmes and maybe also the systemic impact with a good Confidence Ratio.

The goal of Ongoing Evaluations is in part to measure or evaluate outcomes attributable to real public programmes realisations in known contexts (Realistic Evaluation with the model Context-Mechanisms-Outcomes). These outcomes have to be summarized in all their various dimensions and not just financially. Indicators used in the monitoring and evaluation of these outcomes are therefore numerous and very often simplified by the construction of a synthetic one.

Recall that we are interested in the systemic impact, i.e. direct and indirect effects of the program and in the temporal evolution of the impact index, i.e. the evolution of a synthetic indicator obtained by using a composite law (either fuzzy integrals or algorithms) to compose with the diffrent outcome indicators.

For each given moment over the life of a programme, impact is defined by a single combination of consequences outcoming from its implementations (which have to be differenciated from the planned or intended actions). Implementations and outcomes constitute the "value chain" of the deliverables of the activity of public services.

The causal or logic based value chain is drawn up by linking (with a causal link) implementations and outcomes produced by the public policy. It is an essential step of the methodology. To do this, evaluators use a IOI Diagram (Implementations-Outcomes-Impacts). This IOI is a well-known tool for modelling the process of monitoring and evaluating. It allows us to visualize the logic of service interventions with a diagram composed of action plans, implementations (which are engendered with human and financial costs), outcomes and impacts.

Figure 1 : IOI Diagram (Implementation-Outcomes-Impact)

Implementations and their outcomes can be closely evaluated. In other words, it is possible to define one or several indicators which illustrate and measure the nature and intensity of the

outcomes. These indicators are generally stored on a scoreboard and constitute the basic information in different value dimensions that we need to build a synthetic index i.e. as an impact is a composition of outcomes, the index of this impact is by definition a composition of indicators of these outcomes. To measure the impact, evaluators usually decide to apply a composition law on the outcomes indicators. The impact is the synthesis of the multidimensional outcomes in the long-term.

2. Multicriteria analysis methods: comparing options or programmes with one unique synthesis criterion or operate a multicriteria outranking of temporal and multidimensional outcomes i.e. profiles.

2.1. The decision matrix

In multicriteria literature, a decision aid helps decision-makers to make a choice between different options. All of these methods rely, in the first instance, upon the creation of a decision matrix. The decision matrix describes in rows, the different options; and in columns, the different criteria, which allows for the assessment of these options. In this way, we can look up grades for each option relative to each criterion.

2.2. Single synthetic criterion or comparison of profiles

Methods of multicriteria analysis generally fall into two distinct operational approaches in order to obtain a ranking between two options: single synthetic criterion and outranking of profiles. The first approach synthesises the decision matrix by building a global grade for each option and then comparing these global grades; the second makes a pair wise comparison,

criterion by criterion, comparing them over a long time period and postponing the aggregative judgement for as long as possible.

The most renowned method of the "unique synthesis criterion" approach requires the use of a weighted arithmetic average. This method is particularly favoured by international organisms as the PNUD. However, there are many advantages that speak in favour of the second approach: the comparison of profiles.

Some of these advantages are linked to the composition law used, which is a voting rule. Each criterion can be considered as a judge (Roy and Bouyssou, 1993), whereby each judge supports one criterion. Each judge (or criterion) then votes for his preferred option. Thus, outranking methods are an election process for the best option as considered by different judges. To specify the multicriteria social choice, judges have to deliberate on the outranking of the options. In this way, we use a voting rule often linked to the notion of majority rule. Using this process allows us to reasonably aggregate the different preferences expressed by different criteria (or judges).The election process itself consists in putting forward a simple hypothesis of preference on the comparison between two options. For example where **a** and **b** are two options; an hypothesis of preference can be "option **b** is at least as good as option **a**".

Let us say that some judges (or criteria) have classified option **b** as being at least as good as option **a**. This set of judges constitutes a coalition. It is now possible to establish this as a voting rule when conducting the synthesis. This type of voting rule is qualified as a 'majority': any option **b** will now actually be preferred over option **a,** if the coalition of judges respond to this specific voting rule. One has to bear in mind that the value used for the indifference threshold when using this type of voting rule can have a significant effect upon the outcomes. Moreover, if no weighting is defined, then the number of judges is equal to the

number of criteria on which the options are judged. If a weighting function is applied, this has the effect of artificially increasing the number of judges of each criterion.

Methods based on voting rules allow for the addition of veto thresholds. The role of these veto thresholds is to limit the over compensation between criteria. If an option graded by a criterion falls below the veto threshold, then this criterion overrides the vote.

3. Using comparisons of profiles to monitor an Ongoing Evaluation

Our proposition consists of establishing a transformation of the "impact matrix" of an Ongoing Evaluation into a "decision matrix" that monitors profiles of outcomes directly attributable to real implementations. In other words, our proposition considers the different periods of evaluation of outcomes of action plans as so many "options". The temporal evaluation uses static option comparisons, which give an a-temporal specificity to the comparisons of options. The same programme remains unchanged, but its outcomes, ex-post, are considered as options.

To appreciate the possible improvements of the different outcomes of the policy, we use profile comparisons over different periods. In the same way that ranking options outcomes in a composition law, ranking outcome profiles over each period outcomes in a composition law that assesses indicators by each outcome.

The effectiveness of this lemma can be demonstrated as follows. Let us consider a policy that has been defined by a set of actions. For reasons of monitoring and evaluation, outcomes produced by this action plan are tracked over several years. The goal is to identify whether the evolution of the policy, year on year, is good, not so good or bad. As mentioned

earlier, rows of the decision matrix contain grades of one option on each criterion in a decision aid context. In the temporal monitoring of a policy case, these rows will correspond to an action plan a, over different years denoted by $t+i$; let $A=\{a_t,a_{t+1},...,a_{t+n}\}$ be the set of action plans. Let us suppose that some outcomes have been chosen to assess these action plans. Let us assume that a set of indicators directly attributable to these outcomes have been defined for evaluating each period. Let n be the number of indicators with $n \geq 2$. Let g_j be an indicator for the period j where j=1 to n. Let $F=\{g_1,...,g_n\}$ be the set of indicators. The outcoming table contains the grades of the action plan denoted $g_j(a_{t+i})$ of the year considered.

Figure 2 : Decision Matrix

The implementation of a composition law on these grades allows for the determination as to whether action plan a_{t+i} has a better profile than action plan a_{t+i+1}. In other words, it can be determined if the evolution of the policy or program has been positive or negative.

It is now possible to apply the methodology, as defined by international organisations, upon this decision matrix in a monitoring context. After scales have been normalized, each period will be graded by the application of a weighted arithmetic average on the marginal assessment for each outcome.

$$N(a_i)=\frac{w_1 g_1(a_i)+w_2 g_2(a_i)+...+w_n g_n(a_i)}{\sum_{k=1}^{n} w_k}$$

These different grades can be compared to allow for the ranking of different periods and to obtain a formalization of the evolution of an achievement of a policy or programme. As with all decision criteria, the arithmetic average does not merely introduce information into

preferences, it also models them(Bouyssou and al, 2006) . Using an arithmetic average notifies the decision maker that a "full compensation" exists between the outcomes: one unit of a outcome completely offsets the unit of another outcome. Moreover, the weighted function (which is often presented as the outcome of a strong policy choice) depends on the kind of normalization chosen and on the evolution of the data in time. Moreover, whilst the application of this method is simple, defining its parameters is not so easy and reveals strong constraints that subject the decision maker into adapting his preferences to that of the model.

3.1. Using ELECTRE methods in programme monitoring

To establish the monitoring of action plans with inter-temporal outcome profiles, we suggest referring to other decision criteria that call for voting systems. These methods are well known under the term of "outranking methods". There are three methods that can be used: ELECTRE II, ELECTRE III and ELECTRE IV (Roy and al, 1993). These three methods are most often used (André and Roy, 2007) because their algorithms have already been created and are present in several computer applications.

How to make a choice between these three methods? As with the weighted arithmetic average, the technical choices are more often strongly political ones. Remember that the principle of these methods is to fix a voting rule to determine the social choice: that period **b** will actually be preferred to the period **a** if the coalition of judges (or outcomes) consider "period **b** is at least as good as the period **a**" in response to a specific voting rule.

1. When the goal of an actual policy is to reach many objectives, they evolve in systems with several dimensions which are not always substitutable. For example, the concept of sustainable development itself lays down a parallel evolution of economic, environmental and social dimensions in order to reach a durable objective. To accept a perfect substitutability between these dimensions can be considered as a renunciation of the

concept itself. When a decision-maker wishes to strictly apply this principle, one must find a solution that avoids infinite compensation. The three methods suggest using veto thresholds and their role is to limit the compensation effect. Let v_j denote the veto threshold of the indicator g_j. If the difference between the grade of period a and the grade of period a' denoted $\Omega_j(a',a)$ (Dias and al, 2010) is higher than v_j such as $\Omega_j(a',a) \geq v_j$, then the criterion overrides the vote and disallows the implementation of a probable preference.

2. The definition of weights can sometimes pose ethical or policy questions for decisional authorities. Increasing the weight on one criterion (or indicator) artificially increases the weight of the preferences of this indicator. Not implementing weights is a possible solution when interest conflicts occur. Should this be the case, the ELECTRE IV method can be used.

3. Denote g_j an indicator and a and a' two periods. Two cases are possible. In the first case, the "preference relation" between periods a and a' expressed by g_j is a binary relation (a is preferred to a' denoted aP_ja', or a' is preferred to a denoted $a'P_ja$). If the set of indicators have this type of preference, then ELECTRE II has to be used.

4. In the second case the outcome indicators reveal a degree of uncertainty that is somewhat important. In response to this slight degree of uncertainty thrown up by the indicators, it might be preferable to only establish a preference if the differences between grades on each indicator have reached a certain level. In this way official authorities can make the proviso that a differential of preferences between two periods a and a' is too marginal to constitute the revelation of a preference. Let $\Omega_j(a',a)$ denote the advantage of the period

a compared to the period a' (Dias and al, 2010). This advantage can be defined as follows:

$$\Omega_j(a,a') = \begin{cases} g_j(a) - g_j(a') & if \quad g_j \quad is \quad to \quad be \quad maximized \\ g_j(a') - g_j(a) & if \quad g_j \quad is \quad to \quad be \quad minimized \end{cases}$$

To proceed from a binary system of preferences to a progressive one, two thresholds are usually employed: an indifference threshold denoted q_j, and a preference threshold denoted p_j, such that $p_j \geq q_j \geq 0$. These thresholds are associated with each indicator g_j. Their value can be interpreted as follows:

- $\left|\Omega_j(a',a)\right| \leq q_j$ represents a non-significant advantage of the period a over the period a'. This means that the action plan at period a' has had no more effect than at period a, as measured by indicator g_j. In other words, the relation between the period a and the period a' is indifferent, denoted by aI_ja'.

- $\left|\Omega_j(a',a)\right| > p_j$ represents a significant advantage of the period a over the period a'. This means that the action plan at the period a' has had more effect than at the period a, as measured by indicator g_j. In other words, the relation between the period a and the period a' is preferred, denoted by aP_ja'.

- $q_j < \Omega_j(a',a) \leq p_j$ represents an ambiguous outcome. That is to say, the advantage of the action plan over the period a is not significant enough to conclude a relation of strict preference in favour of a, and too significant to conclude a relation of indifference between the periods a and a'. This means that the decision-maker cannot decide between preference and indifference. This situation is defined by a

weak preference in favour of the period a. This specific relation of preference is denoted by $aQ_j a'$.

5. Finally, in an example, let us consider wages as an indicator denoted by g_s. If the policy choice has an outcome indicator with a semantic type "Wages have increased", then constants will be sufficient to define thresholds. Now let us consider the policy choice as an impact with a semantic type "evolution of the economic dimension of the welfare". If the based wage $g_s(a_0)$ is low, a little positive variation of the wage between the two periods a_0 and a_1 could be considered sufficient enough to conclude a preferred relation ($a_1 P_j a_0$). Contrarily, the higher the based wage, the more positive the variation will have to be to conclude in an improvement of the economic welfare. In practice, outranking methods make the implementation of variable thresholds possible. When two periods a_0 and a_1 are compared, the higher the based wage $g_s(a_0)$ is, the greater the difference $g_s(a_1) - g_s(a_0)$ will have to be to obtain a preference such as $a_1 P_j a_0$. Bear in mind that the concordance index is equal to zero if, and only if, the difference between the grades of a_1 and a_0 are greater than the concordance threshold such as $c_s(a_0, a_1) = 0 \Leftrightarrow p_s < g_s(a_1) - g_j(a_0)$, therefore as the value of $g_s(a_0)$ increases, the higher the concordance threshold p_s will have to be. We can determine this with the following function $f = p_s(g_s(a))$ such as $\frac{df}{dg_j(a)} > 0$.

Once the method has been chosen and the parameters have been defined, the algorithm is run on the decision matrix (Maystre and al, 1994)).

3.2. Using the ELECTRE TRI method

Whilst awaiting the completion of a policy implementation, a decision-maker can plan intermediate objectives in order to be sure that it reaches its long-term objectives. A decision-maker may wish to build an indicator that is capable of determining if the public policy has an evolution that is complying with predictions. An Index can be described as a division between the measure of a synthetic index and its goal. This type of measure is easier when both the synthetic index and the goal are numbers. However, in several cases, it is not possible to find a relevant composition law to build a global grade for the impact being considered, and so define a goal. Instead of fixing just one goal for the impact, one solution is to fix one goal for each outcome indicator and to compare, by outcome, the level of each indicator and the level of the goal with which it is associated and then to compare the outcome profiles.

Once more, the implementation of outranking methods presents a good alternative to the problem affecting the options of categories. Each category C_t is delimited by two targeted profiles arbitrarily defined by politics and denoted by h_t and h_{t+1}. These profiles are defined by a vector of goals denoted $V_{t+1} = (g_1(h_{t+1});...;g_n(h_{t+1}))$. In the ELECTRE TRI method, the outranking relation is build in order to enable the comparison of an alternative a to a profile h_t. This outranking relation is build through the following steps:

- compute the partial concordance indices $c_j(a,h_t)$ and $c_j(h_t,a)$;

$$\begin{cases} if \rightarrow g_j(a) \le g_j(h_t) - p_j(h_t), then \rightarrow c_j(a,h_t) = 0 \\ if \rightarrow g_j(h_t) - p_j(h_t) < g_j(a) \le g_j(h_t) - q_j(h_t), then \rightarrow c_j(a,h_t) = \dfrac{\left[g_j(a) - g_j(h_t) + p_j(h_t)\right]}{\left[p_j(h_t) - q_j(h_t)\right]} \\ if \rightarrow g_j(h_t) - q_j(h_t) < g_j(a), then \rightarrow c_j(a,h_t) = 1 \end{cases}$$

$$\begin{cases} if \rightarrow g_j(a) \geq g_j(h_t) + p_j(h_t), then \rightarrow c_j(h_t, a) = 0 \\ if \rightarrow g_j(h_t) + q_j(h_t) \leq g_j(a) < g_j(h_t) + p_j(h_t), then \rightarrow c_j(h_t, a) = \dfrac{\left[g_j(h_t) - g_j(a) + p_j(h_t)\right]}{\left[p_j(h_t) - q_j(h_t)\right]} \\ if \rightarrow g_j(a) < q_j(h_t) + q_j(h_t), then \rightarrow c_j(h_t, a) = 1 \end{cases}$$

- compute the global concordance indices $c(a, h_t)$;

$$c(a, h_t) = \frac{\sum_{j \in F} w_j . c_j(a, h_t)}{\sum_{j \in F} w_j}$$

$$c(h_t, a) = \frac{\sum_{j \in F} w_j . c_j(h_t, a)}{\sum_{j \in F} w_j}$$

- compute the partial discordance indices $d_j(a, h_t)$ and $d_j(h_t, a)$;

$$\begin{cases} if \rightarrow g_j(a) > g_j(h_t) - p_j(h_t), then \rightarrow d_j(a, h_t) = 0 \\ if \rightarrow g_j(h_t) - v_j(h_t) < g_j(a) \leq g_j(h_t) - p_j(h_t), then \rightarrow d_j(a, h_t) = \dfrac{\left[g_j(h_t) - g_j(a) - p_j(h_t)\right]}{\left[v_j(h_t) - p_j(h_t)\right]} \\ if \rightarrow g_j(h_t) - v_j(h_t) < g_j(a), then \rightarrow d_j(a, h_t) = 1 \end{cases}$$

$$\begin{cases} if \rightarrow g_j(a) \leq g_j(h_t) + p_j(h_t), then \rightarrow d_j(h_t, a) = 0 \\ if \rightarrow g_j(h_t) + p_j(h_t) < g_j(a) \leq g_j(h_t) + v_j(h_t), then \rightarrow d_j(h_t, a) = \dfrac{\left[g_j(a) - g_j(h_t) - p_j(h_t)\right]}{\left[v_j(h_t) - p_j(h_t)\right]} \\ if \rightarrow g_j(a) > g_j(h_t) + v_j(h_t), then \rightarrow d_j(h_t, a) = 1 \end{cases}$$

- compute the fuzzy outranking relation grounded on the credibility indices $\sigma(a, h_t)$;

$$\sigma(a, h_t) = c(a, h_t) . \prod_{j \in \overline{F}} \frac{1 - d_j(a, h_t)}{1 - c(a, h_t)} \text{ with } \overline{F} = \left\{ j \in F / d_j(a, h_t) > c(a, h_t) \right\}$$

- and determine a λ-cut of the fuzzy relation in order to obtain a crisp outranking relation.

The cutting level λ is consider as the smallest value of the credibility index compatible with the assertion "a outrank h_t", i.e., $\sigma(a,h_t) \geq \lambda \Rightarrow aSh_t$. We define the binary relations ϕ (preference), I (indifference), R (incomparability) as following: $aIh_t \Leftrightarrow aSh_t$ and h_tSa; $a \phi h_t \Leftrightarrow aSh_t$ and not h_tSa; $a \pi h_t \Leftrightarrow$ not aSh_t and h_tSa ; $aRh_t \Leftrightarrow$ not aSh_t and not h_tSa.

Figure 3 : Decision matrix

At the commencement of the policy $t = 0$, if a_1 is judged by the criteria at least as good as a_0, then the policy is considered as being efficient. This means that the targeted profile h_1 is not arbitrarily defined but is the vector of the grades of the action plan of period a_0 and we can denote that with $V_1 = (g_1(a_0);...;g_n(a_0))$. In other words, a_1 will have to belong to the category C_1. If a_1 belongs to the category C_0, then the policy will not reach its goals. In the same way for the period a_2, if a_2 is judged by criteria at least as good as the targeted profile h_2. That means that the targeted profile h_2 is a vector of grades defined as follows: $V_2 = (g_1(h_2);...;g_n(h_2))$. If a_2 belongs to the category C_2 delimited by h_1 and h_2, then the policy will be considered as being efficient. On the other hand, if a_2 belongs to the category C_1, then the policy will not reach its goals.

Figure 4 : Schematization of the decision matrix with ELECTRE TRI

The comparison of outcome profiles becomes a navigational tool whereby long-term objectives can be reached by continually monitoring the evolution of the policy.

Figure 5 : Links between outranking methods and synthetic indicators

Conclusion

Usually, in the evaluation of public policies and programmes, the goal is to measure outcomes in different dimensions. These outcomes are considered (by convention or after a causality test) as directly attributable to realisations and implementation of the programmes. Evaluators often want a formula for assessment of a more systemic impact (direct and indirect effects of programmes). The impact index is synthetic and follows composition laws that are built on the basis of arithmetical mean or very complex fuzzy integrals.

We have proposed to exclude the calculation of a synthetic index. The different outcomes of each year constitute a profil of results. For comparing the profiles and their temporal evalution, a decision matrix is used in conjunction with multicriteria analyses. We can then use an outranking method on this decision matrix. In this way, we have no need to use integrals and synthetic index. Instead we can happily use a pair wise comparison to obtain a ranking. We have shown that it is possible with ELECTRE software assistance to implement these methods on monitoring and evaluation problems when the need takes the shape of making a synthetic scoreboard. Our future work will concentrate on real life applications.

Bibliography

André, Stephane, and Roy Bernard. "Conception et mise en place d'un outil d'évaluation de la performance environnementale- Le cas des raffineries de la société TOTAL." *Journal of Decision Systems*, no.16(3) (2007): 335-367.

Baslé, Maurice. *Economie, Conseil et Gestion publique. Suivi et évaluation de politiques publiques et de programmes*. Economica, 2008.

Bouyssou, Denis, Marchant Thierry, Pirlot Marc, Alexis Tsoukias, Philippe Vincke. *Evaluation and Decision Models with Multiple Criteria: Stepping Stones for the Analyst*. Birkhäuser: 2006. 445

Dias, Juscelino Almeida, Figueira José Rui, Roy Bernard. "*ELECTRE TRI-C: A Multiple Criteria Sorting Method Based on Central Reference Actions*". European Journal of Operational Research, no. 204 (2010): 565-580.

Daniel, Kaufmann, Kraay Aart, and Mastruzzi Massimo. "Governance Matters VII: Aggregate and Individual Governance Indicators." *World Bank Policy Research Working Paper*, no. 4654 (2008). <http://ssrn.com/abstract=1148386>

Guitouni, Adel, and Martel Jean-Marc. "Tentative guidelines to help choosing an appropriate MCDA method." *European Journal of Operational Research*, no. 109 (1998): 501-521.

Marchant, Thierry, Bouyssou Denis, Perny Patrice, Pirlot Marc, Tsoukias Alexander and Vincke Philipp. " Les indicateurs en perspective." *Preprints of the Service de mathématiques de la gestion* 2003-09 (2003).

Maystre, Lucien Yves, Pictet Jacques, and Simos Jean. *Méthodes multicritères ELECTRE, Description, Conseils pratiques et cas d'application à la gestion environnementale*. Presses Polytechniques et Universitaires Romandes: 1994.

Perret, Bernard. " Indicateurs sociaux, état des lieux et perspectives." *Rapport au Conseil de l'Emploi, des Revenus et de la Cohésion Sociale*, Paris (2002).

Roy, Bernard, and Bouyssou Denis. *Aide multicritère à la décision : méthodes et cas*. Economica: 1993.

Roy, Bernard, ed. "Partial preference analysis and decision-aid: The fuzzy outranking relation concept." *Conflicting objectives in decisions.* Bell D.E., Keeney R.L., Raiffa H., Wiley: New York, 1978.

Yoon, K. Paul and Hwang Ching-Lai. *Multiple attribute decision making: An introduction*. California: Sage University Press, 1994.