Multicriteria analysis for environmental assessment of agri-environment schemes: how to use partial information from Mid-Term Evaluations?

Davide Viaggi[1], John Anthony Finn[2], Isabelle Kurz[2,3] and Fabio Bartolini[1]

Abstract

This paper addresses how environmental indicators and multicriteria methodologies can support evaluation of the environmental performance of EU agri-environment schemes (AESs). Prominent evaluation issues and problems concerning AES are discussed, with a focus on availability of information. With these issues considered, a multicriteria methodology was devised and applied to estimate the environmental effectiveness of an AES in each of two study areas: Ireland and Emilia-Romagna (Italy). The evaluation is based on information from the mid-term evaluation of the Rural Development Programmes. The results suggest that the AESs only partially achieved their objectives. This interpretation is tentative, largely due to the scarcity of quantitative data that related to effectiveness, the lack of quantitative target levels for objectives, and difficulties in determining the relative importance of different environmental objectives.

Key words: *Agri-environment schemes, mid-term evaluation, multifunctional agriculture, indicators, multicriteria analysis.*

Background and objectives

As part of the Common Agricultural Policy, agri-environment schemes (AESs) provide payments to farmers in exchange for the provision of environmental goods. Agri-environment schemes were implemented as a pan-European policy mechanism under reg. CE 2078/92, and have proceeded through several 5- to 7-year policy cycles since then. By 2002, about 25% of all agricultural land in EU was under AES agreements (EEA, 2005). Payments for different measures ranged from about €20 per hectare to over €750 per hectare, and total annual expenditure on EU AESs since 2000 has consistently exceeded €2 billion per annum. Given the scale of expenditure, evaluation of policy action is required to continually improve performance, ensure value-for-money for taxpayers, and avoid accusations of trade distortion.

In advance of the 1999 iteration of AESs, considerable effort was made to produce a set of common questions for the comparative evaluation of Rural Development Programmes (RDP), including AES, across Europe (European Commission, 2000; 2002a;

[1] (Corresponding author)Department of Agricultural Economics and Engineering (DEIAGRA), University of Bologna (Italy)
[2] Teagasc Environment Research Centre, Johnstown Castle, Wexford, Ireland
[3] Teagasc Rural Economy Research Centre, Athenry, Co. Galway, Ireland

2002b). Based on this framework, many countries/regions in Europe produced their Mid-Term Evaluations (MTEs) in 2003/2004. Some of the MTEs were also updated in 2005 and some regions performed an ex post evaluation in 2006/2007. MTEs were intended to produce an intermediate evaluation half way during the period, to allow adjustments. However, as the design and implementation of programs for the next programming period started before the final evaluations were available, MTE become one of the main information supports for decisions on the design of forthcoming schemes. In some cases, MTE documents have been used by Member States to support the ex ante analysis of the set of agri-environment measures developed for the 2007-2013 period.

Despite the emphasis given to evaluation procedures, the process of policy evaluation of AESs up to now has been considered unsatisfactory to a large extent. Despite producing a large amount of qualitative and technically detailed information for local decision making about specific measures, the information produced only allowed little formalised judgement, and was weakly comparable across areas (Oréade-Brèche, 2005). In addition, the outcome of the evaluations is still difficult to interpret for wider evaluation exercises due to weakly defined objectives and incomplete data availability. In response to these difficulties, the European Commission (2006) defined a new monitoring and evaluation system to support the implementation of Rural Development Plans (RDP) for the programming period 2007-2013 (reg. CE 1698/2005). However, the production of a consistent evaluation that collates across a prescribed system of information collection is still unresolved.

The objective of this paper is to compare alternative approaches to ex post evaluation of AESs, focusing on environmental assessment and taking into account limitations in the availability of information. Research on AES evaluation has proposed a number of tools and methodologies. However, few of these tools may help in organising ex post information and are not suitable when only a very partial amount of information is available. The paper builds on information drawn from the MTE documents produced during the intermediate round of evaluation of the AESs for the programming period 2000-2006 (in response to the common questions set by the European Commission) and focuses on an example in which Ireland and Emilia-Romagna are compared. In order to develop such comparison, some alternative multicrieria approaches are devised and applied to the two case studies. The paper will proceed through the following outline: overview of AESs evaluation problems (section 2); suitability of multicriteria methodologies and description of the methodology adopted (section 3); results (section 4); and discussion (section 5).

Overview of indicators and MCA in connection to AES

Simply put, evaluation is a process that aims to assess performance and thereby identifies strengths and corrects weaknesses. The ability to assess ‘performance’ implies some a priori definition of what is a poor, satisfactory, good or excellent performance level. Thus, we consider ‘effectiveness’ to be a measure of how well the actual performance matches the expected performance level (Finn, 2008).

These approaches also underpin the evaluation of the environmental effectiveness of AESs. Ideally, an AES will have clearly stated environmental objectives, for which there are specific, measurable environmental targets to be achieved. The aim of monitoring is to collect information on the actual environmental performance, which can then be compared with the original, expected environmental targets. The comparison of collected data with quantitative targets then forms the basis of the objective decision-making that is the purpose of an evaluation. Although monitoring involves the collection of data, evaluation uses the data to interpret the effectiveness of the scheme and make decisions on the basis of evidence. In this way, the evaluation process can:

- identify the extent to which the scheme objectives are being fulfilled, and;
- identify any changes required to bridge the gap between policy aims (environmental targets) and policy outcomes (actual environmental performance).

Thus, the evaluation process can confirm that elements of a scheme are effective and, where necessary, recommend amendments to improve effectiveness. As such, the agri-environmental evaluation is an iterative process that facilitates the flexibility required for continual improvement of agri-environment schemes.

The implementing authorities of AES in EU countries are making an important effort to collect information about the effectiveness of AES (Oréade-Brèche, 2005). However, such information usually does not satisfactorily underpin the quantitative analysis required for evaluation procedures. The main challenges to satisfactory evaluation include: a) adequate definition of the objectives to be achieved; b) adequate measurement of the effects of the schemes; c) the aggregation of different indicators in an analysis, and; d) the comparability of evaluation results across different schemes and locations. We briefly address each of these in turn. First, scheme objectives may lack clarity about the environmental objectives to be addressed and often lack the specific, measurable environmental targets to be achieved. Any attempt to assess the performance of AESs at a European scale is also complicated by the diversity of agri-environmental objectives across Member States, and even among different regions within a Member State. Second, the measurement of scheme performance is often restricted to an incomplete set of indicators (i.e. information may not be collected for all relevant objectives). In addition, information is often limited to uptake indicators such as participation, while measurement of environmental impacts are poor (Court of Auditors, 2000). It is also difficult to disentangle the additional effects of policies from the counterfactual situation in which a policy is not applied. Third, most AESs simultaneously pursue different environmental objectives via different agricultural management prescriptions (Primdahl *et al*, submitted; Purvis *et al*. 2009). Despite this, AES evaluations often lack either a means of aggregation or consideration of trade-offs among different indicators/objectives. Fourth, judgements about the outcome of schemes are rarely based on correct measures of the counterfactual situation. Performance over time is difficult to measure, due to the limited availability of relevant farm-scale environmental data collected in a time series. On the other hand, comparison of different achievements in different study areas is often unreliable due to complex relationships among the contextual environmental and economic features, intended objectives, appropriate incentives as well as willingness to participate and compliance by farmers.

A number of explanations may account for the above challenges to evaluation. First of all, AESs address a number of different environmental issues, which require measurement of a number of different indicators. There is not usually an effective analytical framework with which to use such information to aid interpretation, which would reinforce the importance and collection of such information. In some cases, impacts may be very costly to measure and assess. In such cases, measurement of changed adoption of practices may be the only feasible evaluation of policy effects. Secondly, the evaluation process is recent, at least in a policy-relevant time frame. Environmental effects may be associated with significant lag times, and measurable responses to a policy may take longer to manifest than the 2-3 years usually available for measuring the effects within the time frame of the scheme. Thirdly, each scheme has unique features, and these evolve over time. Such variation over time and space may confound comparisons between earlier and later versions of a scheme, as well as comparison of different schemes. Fourthly, with greater clarity about scheme objectives comes greater accountability. Thus, it is reasonable that decision makers do not like to expose themselves to specific targets any more than is necessary, particularly when effects, and even uptake, are so difficult to predict. In some cases, there was very low uptake of some measures that resulted in low environmental performance, whereas uptake was so unexpectedly high in other cases that selection procedures were required to limit expenditure.

Quality of information from monitoring and evaluation procedures is improving. However, it is reasonable to expect that the aforementioned obstacles and issues will not change substantially in the near future, in spite of the new evaluation framework (European Commission, 2006). Here, we attempt to address a number of specific challenges that face AES evaluation, focusing in particular on a) the formulation of a performance (effectiveness) judgement for individual measures, b) the aggregation of performance concerning the multiple objectives affected by AESs, c) the ways to tackle the lack of complete measures of AES performance. We devise a relatively simple methodology that could help decision makers in assessing the environmental effectiveness of schemes based on available partial information.

The methodology

Reference tools and overview

Multicriteria analysis (MCA) encompasses a number of techniques that aim to evaluate alternative courses of action (e.g. alternative policy design options) on the basis of multiple criteria. A number of multicriteria methodologies have been developed (Saaty, 1980; 2000; Zeleny, 1982; Roy, 1985; Maystre *et al.* 1994) and are reviewed in Guitouni and Martel (1997). The multicriteria approach has been used since the 1980s on a variety of issues, including environmental impact assessment, policy assessment and project evaluation. It is particularly suitable to support participatory decision making, as it allows the comparison of alternatives on the basis of the relevant evaluation criteria, their relevance and their effects on the final results. A theoretical analysis of the application of MCA techniques to agriculture and its relations with environmental issues is presented in Rehman and Romero (1993); and reviewed by Hayashi (2000).

In order to achieve an effective evaluation, three key issues include the definition and measurement of evaluation criteria, the choice of the aggregation procedure and the quantification of weights Hayashi (1998). Agri-environment schemes, and multifunctionality issues in general, appear particularly suitable for the use of multicriteria analysis, as policy performance may be measured through a number of indicators and criteria. The EU has provided a list of indicators for the evaluation of AESs (European Commission, 2000). Common indicators have been widely discussed and may be complemented with locally defined criteria. Nevertheless most of them appear insufficient to quantify the real impact, as they are mostly uptake indicators. Also, no real aggregation procedure is devised to achieve an overall picture of the policy performance, allow comparison or estimate trade-offs among objectives.

MCA techniques may be used both for ex post or ex ante evaluation of AESs, though these two options entail rather different applications, especially where data requirements are concerned (Bartolini *et al*., 2007). Ex post evaluation may involve a number of problems that arise from insufficient data as well as the lack of alternative simulated options for comparison. Alternatively, MCA may be used to compare the implementation of policies in different geographical areas in order to understand their relative effects and possibly to identify factors of success in one case compared to the other.

Despite its potential suitability and the fact that current data collection in AES is structured as a grid of indicators (apparently an ideal starting point for an MCA), MCA is still rarely used for practical purposes in AES evaluations. This may be due to the general complexity of the methodologies adopted in the MCA literature, and due to the difficulties in formalising a very qualitative and information-intensive evaluation process in a stylised MCA framework.

The methodology adopted in this paper is based on the use of MCA as a tool for ex post evaluation of the application of AESs, through comparison of environmental performance of AES policy as implemented in a region of Italy (Emilia Romagna) and the whole of Ireland. The basic information required is drawn from the relevant MTE. The decision to not use information additional to that in the MTE (except, partially, for weights) is intentional, in order to mimic the same information conditions faced by an evaluator/policy maker. The methodology runs in two steps. In a first step, the information included in the MTE has been translated into a score along a quantitative scale of different criteria for effectiveness. In the second step, MCA has been applied to the scores obtained for each criterion derived from the first step, through weighting and aggregation.

From MTE information to an assessment of effectiveness

Despite the definition of a common set of evaluation questions to be answered in the MTEs of the RDPs (see section 2), the approaches towards evaluating the measures of the RDPs vary widely among the member states. The MTE reports differ considerably in the type of information they contain and the level of detail they provide. This adds to the aforementioned difficulties in assessing the comparative effectiveness of agri-environmental schemes among EU member states (see Section 2.1). Given the limitations of the information compiled in the MTEs, the environmental effectiveness of AESs was estimated by assessing whether the agri-environment measures in an area achieved the most important medium-level environmental objectives. The medium-level objectives were derived by re-wording the list of indicators devised as a framework to answer the common evaluation questions (European Commission, 2000) (Table 1).

Note that one objective may be achieved through several measures, where a measure is a prescribed management practice that is expected to achieve one or more environmental objectives. In general, the environmental effectiveness of a measure can be assessed by answering a number of questions (Finn *et al*. 2008, 2009):

- is the measure capable of achieving the stated objective i.e. is there a causal link between the management practice and the achievement of the environmental objective?

- has the measure been implemented properly by institutions and participating farmers?
- for each measure, what minimum participation rate is required to achieve the named objective (desired participation rate), and how does this compare with the actual participation rate?
- what proportion of participants agree to implement the measure, but do not (compliance)?

Table 1 - List of environmental issues represented by medium-level objectives (criteria) grouped by sub-factors. Summarised from Common Evaluation Questionnaire (modified from European Commission, 2000)

VI.1.A. To what extent have natural resources been protected in terms of soil quality, as influenced by agri-environmental measures?

VI.1.A-1. Soil erosion has been reduced

VI.1.A-2. Chemical contamination of soils has been prevented or reduced

VI.1.B. To what extent have natural resources been protected in terms of the quality of ground and surface water, as influenced by agri-environmental measures?

VI.1.B-1. Reduction of agricultural inputs potentially contaminating water

VI.1.B-2. The transport mechanisms (from field surface or root zone to aquifers) for chemicals have been impeded (leaching, run-off, erosion)

VI.1.B-3. Improved quality of surface water and/or groundwater

VI.1.C. To what extent have natural resources been protected (or enhanced) in terms of the quantity of water resources, as influenced by agri-environmental measures?

VI.1.C-1. The utilisation (abstraction) of water for irrigation has been reduced or increase avoided

VI.1.C-2. Water resources protected in terms of quantity

VI.2.A. To what extent has biodiversity (species diversity) been maintained or enhanced thanks to agri-environmental measures through the protection of flora and fauna on farmland?

VI.2.A-1. Reduction of agricultural inputs (or avoided increase) benefiting flora and fauna has been achieved

VI.2.A-2. Crop patterns [types of crops (including associated livestock), crop rotation, cover during critical periods, expanse of fields] benefiting flora and fauna have been maintained or reintroduced

VI.2.A-3. Species in need of protection have been successfully targeted by the supported actions

VI.2.B. To what extent has biodiversity been maintained or enhanced thanks to agri-environmental measures through the conservation of high nature-value farmland habitats, protection or enhancement of environmental infrastructure or the protection of wetland or aquatic habitats adjacent to agricultural land (habitat diversity)

VI.2.B-1. “High nature-value habitats” on farmed land have been conserved

VI.2.B-2. Ecological infrastructure, including field boundaries (hedges…) or non-cultivated patches of farmland with habitat function have been protected or enhanced

VI.2.B-3. Valuable wetland (often uncultivated) or aquatic habitats have been protected from leaching, run-off or sediments originating from adjacent farmland

VI.2.C. To what extent has biodiversity (genetic diversity) been maintained or enhanced thanks to agri-environmental measures through the safeguarding of endangered animal breeds or plant varieties?

VI.2.C-1. Endangered breeds/varieties are conserved

VI.3. To what extent have landscapes been maintained or enhanced by agri-environmental measures?

VI.3-1. The perceptive/cognitive (visual, etc) coherence between the farmland and the natural/biophysical characteristics of the zone has been maintained or enhanced

VI.3-2. The perceptive/cognitive (visual, etc) differentiation (homogeneity/diversity) of farmland has been maintained or enhanced

VI.3-3. The cultural identity of farmland has been maintained or enhanced

A precursory inspection of the MTEs generally revealed an inadequate level of information on many of these points. Most MTEs did, however, quote evidence, or offer an assessment, of the ability of the management prescriptions to achieve stated objectives. MTEs generally gave information on 1) the area actually covered by a measure (the actual participation rate) 2) the area to which the measure is applicable, or 3) participation rates expected by policy makers. For the purposes of this study, either of points 2 or 3 is used to indicate the participation rates required to achieve an objective. (Both these approaches assume that policy-makers are both correct in choosing the environmental issue to be addressed, and in proposing appropriate target participation rates.)

In a first step, the 'performance' and the 'actual participation relative to desired participation' of measures were summarised by applying a decision-making framework, which is described in the following paragraphs. At the same time, however, the major gaps of information were identified and strategies (outside the existing MTE reports) to fill them developed.

The decision-making process began with the list of most important objectives for each geographical area. The agri-environmental measures that contributed to achieving each relevant objective were then identified. The evidence quoted in the relevant MTE was used to rate the performance of each measure, or group of measures, as low, medium or high. Although subjective, these decisions were strongly guided by the information and judgement that was stated or implied in the MTE. However, if quantitative information was available, a standard approach was used across countries. A reduction of substance usage due to an agri-environmental measure, for example, was set in relation to the 'normal' usage. A reduced usage of 0 to 10% was rated low, whereas reductions of > 10% to 25% and of > 25% were rated medium and high, respectively.

A similar process was followed when estimating the actual participation relative to desired participation. The actual participation in a measure was rated low when it amounted to 0 to 40% of the desired participation, medium if it attained > 40% to 70% and high if it reached > 70% of the desired participation.

The 'performance' and 'participation' ratings were then transposed into numeric values, and the product of the two ratings served as an approximate estimate of the effectiveness of a measure, or a group of measures, in achieving the stated objective.

From effectiveness scores to aggregated ranking parameters

Using the scores produced in the previous step, the MCA was applied to each evaluation criteria. Effectiveness was taken as the value of environmental quality function without further considering thresholds or other non-linearities (Finn *et al*. 2008). MCA was performed using the hierarchical aggregation framework illustrated in Figure 1.

The lowest level in Figure 1 is comprised of the criteria (medium level environmental objectives) already described in Table 1. These criteria can be aggregated according to three hierarchical levels: a) by sub-factor, i.e. the next aggregation in 7 environmental components already used in Table 1; b) by factor, i.e. each of the four major environmental components (soil, water, biodiversity, landscape), and; c) overall, as a single score aggregating all factors.

Before aggregation, we checked for possible overlap among criteria and considered excluding some criteria. However, only VI.1.B-3 was identified as potentially causing overlap, and, as a consequence, it was decided to keep the original list of criteria.

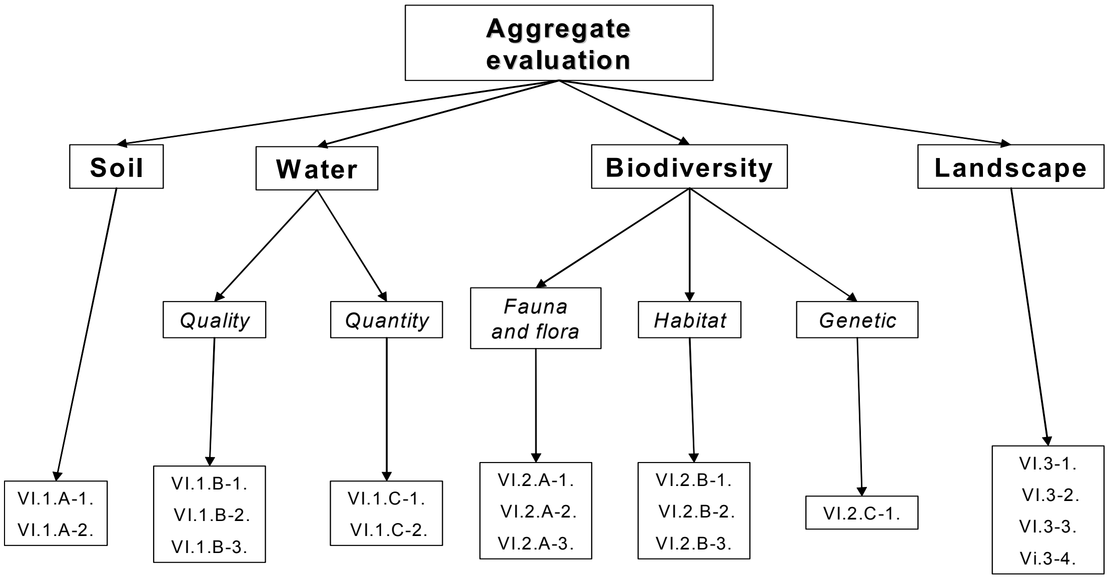

Figure 1 - Hierarchical relationship among the criteria, sub-factors and factors. Source: modified from European Commission (2000).

Aggregation was performed using three different methodologies: a) hierarchical weighted sum omitting criteria (and related weights) for which data were not available; b) hierarchical weighted sum with zero value for criteria for which data were not available, and; c) concordance index for the first level of aggregation.

Options a) and b) were based on a hierarchical weighted sum, where the score was generally given by:

$$u(a_i) = \sum_{i=1}^{k} v_j a_{ij} \tag{1}$$

where:

$u(a_i)$ = utility of the ith alternative (study area/case of application);

v_j = weight of the jth criterion;

a_{ij} = utility value of the ith alternative for the jth criterion.

Weights were quantified by the authors according to an assessment of the relevance of each indicator, where relevance was judged by the degree to which an indicator reflected the proportional priority of environmental objectives. This was based on information drawn from the RDP and MTE, in particular based on the ranking and qualitative consideration of each issue in the documents related to each area. One unusual issue is that weights for specific environmental objectives differed between the two case studies. A weight has been assigned to all relevant criteria, sub-factors and factors.

The main problem with this method is that information is not available for all criteria. In a first option (a), this problem has been solved by re-calculating weights attributing zero to unavailable criteria (those for which it was not possible to calculate a performance score) and redistributing the weights within available criteria in proportion to the original weight. This means in fact achieving an evaluation solely based on quantifiable criteria, ignoring information gaps and allowing the overall result to reflect the performance of the most data-rich measures.

In a second option (b), the methodology has been adjusted by using the original weights, as follows:

$$u(a_i) = \sum_{i=1}^{k} v_j a_{ij} \delta_j \tag{2}$$

where:

$\delta_j = 0$ if the indicator score is not available, 1 if the indicator score is available.

This is the same as attributing a zero score to those indicators for which a performance score could not be quantified. The outcome is a very conservative evaluation, reflecting only the certain achievements for which data are available. As a result, the evaluation can be considered more robust, but also highly dependent on the set of indicators measured in each area (which is normally different) and likely underestimating the total effect of the program.

To address the lack of information for many criteria in a more direct but less extreme manner, a concordance index was also used to compare the two study areas (methodology c above). This index was applied to the first level criteria. The score is based on the comparison across alternatives (areas) for each indicator and attributed using the following rules:

$s_{ii'j} = 1$, if $a_{ij} \geq a_{i'j}$;

$s_{ii'j} = 0{,}5$, if the value of the indicator is missing for i;

$s_{ii'j} = E / E_{\max}$, where E =effectiveness score of the indicator and $E_{\max}$ =maximum effectiveness score attainable (in our case 36), if the value of the indicator is missing for i.

In case of missing information, the simple rationale behind this scoring is to assign: i) equal probability of being better if the value of one indicator is available for none of the two alternatives; ii) a probability value of being better that is proportional to the score of the indicator for the alterative in which it is available, when the other is not available.

This approach complements the previous ones on two grounds. The first is that it is non-compensatory, i.e. high positive difference between two alternatives for some indicators cannot outweigh negative differences. As a result of using this approach, alternatives having a high number of better indicators tend to prevail regardless of the size of the difference between alternatives for each indicator The second is that it accounts for missing indicators by assigning ’probabilistically’ a score to the comparison when one or both the indicators to be compared are not available.

Results

The case study

An explorative case study is illustrated. The case study compares Ireland and Emilia Romagna (Italy). The choice of these two areas was mostly determined by data availability. However, the two areas also reflect very different environmental conditions and very different strategies for policy implementation.

Effectiveness evaluation

Effectiveness criteria and related scores for Ireland show average to good results in terms of contribution to environmental improvement and rather good results in terms of participation related to target (Table 2). The outcome may be to a good extent attributed to a relatively high uniformity of the territory involved and of the measures proposed as well as the relatively simple structure of measures. Also, target levels of objectives were relatively easy attained and sufficient financial budget was available to facilitate such high participation levels.

Table 2 - Results for the first step of evaluation: separate scoring of the performance criteria and participation criteria and their combined final score (Ireland).

Criterion	MTE evidence of performance of measures			Actual participation / desired participation			Final Score	
	Description	Qualitative rating	Numeric value	Description	Qualitative rating	Numeric value	Product	Over-all
VI.1.A-1	no info							
VI.1.A-2	no info							
VI.1.B-1	24 % reduction of P usage	Medium	4	78%	high	6	24	
	Significant improvement in waste storage	High	6	78%	high	6	36	30
	no info							
VI.1.B-2	no reliable evidence							
VI.1.B-3	no reliable evidence							
VI.1.C-1	no info							
VI.1.C-2	no info							
VI.2.A-1	Little reliable evidence	Medium	4	78%	high	6	24	24
VI.2.A-2	no info							
VI.2.A-3	Little reliable evidence	Low	2	78%	high	6	12	12
VI.2.B-1	Little reliable evidence	Medium	4	78%	high	6	24	24
VI.2.B-2		High	6	78%	high	6	36	36
VI.2.B-3	No info							
VI.2.C-1		High	6	substantial decrease	low	2	12	12
VI.3-1	no info							
VI.3-2	no info							
VI.3-3	no info							
VI.3-4	6 504 features identified, 2 128 new.	High	6	78%	high	6	36	36

The case of Emilia Romagna is more complex, due to the higher variety of measures and environmental issues addressed (Table 3). The contribution to environmental improvement may be generally considered as good. However, no clear targets were set at the beginning, which makes it difficult to conduct an ex post assessment. Target levels have been estimated as the amount of land that could have been potentially addressed by each measure. However, this may well overestimate the appropriate target, as it could be excessively optimistic with respect to the available budget and the compatibility among objectives. In fact, the results from Emilia Romagna are probably negatively affected by an insufficient budget to address the size of the environmental issues to be dealt with. In practice, there were budget limits for each of the different environmental issues or measures, and this could result in participation being the limiting factor for the environmental effectiveness of specific measures.

Table 3 - Results for the first step of evaluation: separate scoring of the performance criteria and participation criteria and their combined final score (Emilia Romagna).

Criterion	MTE evidence of performance of measures			Actual participation / desired participation			Final Score	
	Description	Qualitative rating	Numeric value	Description	Qualitative rating	Numeric value	Product	Overall
VI.1.A-1	Increase in minimum or no tillage	High	6	58 894 ha,no targets	low	2	12	12
	Increase in cover crops							
	Increase in organic matter of soil							
	Increase of land use with low potential for erosion.	High	6	9 412 ha, no targets	low	2	12	
VI.1.A-2	Reduced usage of plant protection products.	High	6	66 309 ha, no targets	medium	4	24	24
	Reduced usage of chemical and organic fertilisers.							
VI.1.B-1	Reduced usage of chemical fertilisers: 31% average reduction of N usage, 62% average reduction of P usage.	High	6	32492 ha, no targets	low	2	12	12
	Significant reduction in usage of plant protection products.	High						
	Reduced usage of chemical fertiliser: 93% average reduction of N usage, 39% average reduction of P usage.	High	6	26402 ha, no targets	low	2	12	
	Significant reduction in usage of plant protection products.	High						
	Reduced usage of chemical fertilisers.	High	6	7415 ha, no targets	low	2	12	
	Reduced usage of plant protection products.							
	Reduced usage of organic fertiliser.							
	Increased areas of low input crops.							
	Reduced usage of plant protection products.							
VI.1.B-2	Increase of areas with land cover to impede contaminant losses to water.	High	6	8536 ha, no targets	low	2	12	10
	Increase of areas with features to impede contaminant losses to water.	Medium	4	61839 ha, no targets.	low	2	8	
VI.1.B-3	No info							
VI.1.C-1	Increase of areas with reduced irrigation	medium	4	12645 ha, no targets.	low	4	16	13
	Increase of non-irrigated area.	high	5	686 ha, no targets.	low	2	10	
VI.1.C-2	No info							
VI.2.A-1	Reduced usage of plant protection products.	High	6	53303 ha, no targets.	low	2	12	12
	Reduced usage of fertilisers.							
	Avoidance of inputs during critical periods.							
VI.2.A-2	Increase of areas with crop patterns benefiting flora and fauna.	medium	4	53303 ha, no targets.	low	2	8	8
VI.2.A-3	Field work showing benefit of measures for birds.	high	2	28900 ha, no targets.	low	2	4	4
VI.2.B-1	No info							
VI.2.B-2	No info							
VI.2.B-3	No info							
VI.2.C-1	No info							
VI.3-1	No info							
VI.3-2	No info							
VI.3-3	No info							
VI.3-4	No info							

Multicriteria analysis

In the first stage, MCA has been carried out by excluding the criteria for which no data was available and deriving multicriteria indices based on the set of remaining criteria. Results for Ireland showed a relatively low number of criteria, but a good correspondence between the weights assigned to criteria and the availability of information on that indicator (Table 4). In the intermediate (sub-factor) level, however, the results are better for landscape and water, while the highest weight was attributed to biodiversity, which mostly affected the final result.

Table 4 - Results of the MCA with omission of criteria with missing scores (Ireland)

		A) Aggregation by sub-factor					B) Aggregation by factor			C) Overall score		
Criterion	**Score**	**Weight**	**Weight r corrected**	**Sub-Factor**	**Score**	**Weight r**	**Weight corrected**	**Factor**	**Score**	**Weight**	**Weight corrected**	**Final Score**
VI.1.A-1		1,00	0,00	Soil	0,0	1,00	1,00	Soil	0,0	0,10	0,00	**26,6**
VI.1.A-2		0,00	0,00									
VI.1.B-1	30	0,38	1,00	Water quality	30,0	1,00	1,00	Water	30,0	0,26	0,29	
VI.1.B-2		0,25	0,00									
VI.1.B-3		0,38	0,00									
VI.1.C-1		0,00	0,00	Water quantity	0,0	0,00	0,00					
VI.1.C-2		0,00	0,00									
VI.2.A-1	24	0,43	0,50	Biodiversity (flora and fauna)	18,0	0,35	0,35	Biodiversity	20,4	0,44	0,49	
VI.2.A-2		0,14	0,00									
VI.2.A-3	12	0,43	0,50									
VI.2.B-1	24	0,43	0,50	Biodiversity (habitat)	30,0	0,35	0,35					
VI.2.B-2	36	0,43	0,50									
VI.2.B-3		0,14	0,00									
VI.2.C-1	12	1,00	1,00	Biodiversity (genetic diversity)	12,0	0,30	0,30					
VI.3-1		0,25	0,00	Landscape	36,0	1,00	1,00	Landscape	36,0	0,20	0,22	
VI.3-2		0,00	0,00									
VI.3-3		0,50	0,00									
VI.3-4	36	0,25	1,00									

The first relevant issue for Emilia Romagna is the number of potentially relevant criteria that have not been quantified at this stage, with respect to the more distributed relevance (weights) of environmental issues (Table 5). The analysis indicates a very different policy profile, as Emilia Romagna is more oriented towards soil and water conservation than Ireland. Also, due to the low level of effectiveness attributed to most of the criteria, Emilia Romagna scores a very low overall result.

Table 5 - Results of the MCA with omission of criteria with missing scores (Emilia Romagna)

Criterion	Score	A) Aggregation by sub-factor: Weight	Weight r corrected	Sub-Factor	Score	Weight r	Weight corrected	B) Aggregation by factor: Factor	Score	Weight	Weight corrected	C) Overall score: Final Score
VI.1.A-1	12	0,50	0,50	Soil	18,0	1,00	1,00	Soil	18,0	0,30	0,35	13,1
VI.1.A-2	24	0,50	0,50									
VI.1.B-1	12	0,43	0,75	Water quality	11,5	0,54	0,54	Water	12,2	0,32	0,38	
VI.1.B-2	10	0,14	0,25									
VI.1.B-3		0,43	0,00									
VI.1.C-1	13	0,50	1,00	Water quantity	13,0	0,46	0,46					
VI.1.C-2		0,50	0,00									
VI.2.A-1	12	0,33	0,33	Biodiversity (flora and fauna)	8,0	0,21	1,00	Biodiversity	8,0	0,23	0,27	
VI.2.A-2	8	0,33	0,33									
VI.2.A-3	4	0,33	0,33									
VI.2.B-1		0,20	0,00	Biodiversity (habitat)	0,0	0,36	0,00					
VI.2.B-2		0,40	0,00									
VI.2.B-3		0,40	0,00									
VI.2.C-1		1,00	0,00	Biodiversity (genetic diversity)	0,0	0,43	0,00					
VI.3-1		0,25	0,00	Landscape	0,0	1,00	1,00	Landscape	0,0	0,15	0,00	
VI.3-2		0,25	0,00									
VI.3-3		0,25	0,00									
VI.3-4		0,25	0,00									

The results were changed significantly when a zero value was assigned to the score of those indicators for which information was not available. Although the overall judgement on the comparison between the two areas does not change, the role of different (groups of) criteria changes remarkably. For example biodiversity becomes the best scoring indicator for Ireland, while landscape comes in last place (Table 6).

Table 6 - Results of the MCA with inclusion of criteria with missing scores (Ireland)

Criterion	Score	A) Aggregation by sub-factor: Weight	Sub-Factor	Score	Weight	B) Aggregation by factor: Factor	Score	C) Overall score: Weight	Final Score
VI.1.A-1		1,00	Soil	0,0	1,00	Soil	0,0	0,10	12,7
VI.1.A-2		0,00							
VI.1.B-1	30	0,38	Water quality	11,3	1,00	Water	11,3	0,26	
VI.1.B-2		0,25							
VI.1.B-3		0,38							
VI.1.C-1		0,00	Water quantity	0,0	0,00				
VI.1.C-2		0,00							
VI.2.A-1	24	0,43	Biodiversity (flora and fauna)	15,5	0,35	Biodiversity	18,0	0,44	
VI.2.A-2		0,14							
VI.2.A-3	12	0,43							
VI.2.B-1	24	0,43	Biodiversity (habitat)	25,8	0,35				
VI.2.B-2	36	0,43							
VI.2.B-3		0,14							
VI.2.C-1	12	1,00	Biodiversity (genetic diversity)	12,0	0,30				
VI.3-1		0,25	Landscape	9,0	1,00	Landscape	9,0	0,20	
VI.3-2		0,00							
VI.3-3		0,50							
VI.3-4	36	0,25							

Table 7 - Results of the MCA with inclusion of criteria with missing scores (Emilia Romagna).

		A) Aggregation by sub-factor				B) Aggregation by factor		C) Overall score	
Criterion	Score	Weight	Sub-Factor	Score	Weight	Factor	Score	Weight	Final Score
VI.1.A-1	12	0,50	Soil	18,0	1,00	Soil	18,0	0,30	7,86
VI.1.A-2	24	0,50							
VI.1.B-1	12	0,43	Water quality	6,6	0,54	Water	6,5	0,32	
VI.1.B-2	10	0,14							
VI.1.B-3		0,43							
VI.1.C-1	13	0,50	Water quantity	6,5	0,46				
VI.1.C-2		0,50							
VI.2.A-1	12	0,33	Biodiversity (flora and fauna)	8,0	0,21	Biodiversity	1,7	0,23	
VI.2.A-2	8	0,33							
VI.2.A-3	4	0,33							
VI.2.B-1		0,20	Biodiversity (habitat)	0,0	0,36				
VI.2.B-2		0,40							
VI.2.B-3		0,40							
VI.2.C-1		1,00	Biodiversity (genetic diversity)	0,0	0,43				
VI.3-1		0,25	Landscape	0,0	1,00	Landscape	0,0	0,15	
VI.3-2		0,25							
VI.3-3		0,25							
VI.3-4		0,25							

The same happens in Emilia Romagna, where the result for soil is strengthened with respect to water and biodiversity (Table 7). When comparing the two study areas, the result again appears rather straightforward, as Ireland is better than Emilia Romagna at all levels of aggregation and for all criteria. To a good extent, however, this is due to a lack of information for many criteria that would have complemented and possibly changed the results.

An attempt to understand the possible relevance of additional criteria is performed using the concordance index for the first step of aggregation. This is done in Tables 8, followed, in Table 9 by aggregation of the outcome through the following levels of the hierarchical objective structure. The main relevant differences are more evident in the first stage reported in Table 8, which allows a rough estimate of the role of missing criteria in the comparison across study areas.

Table 8 – Comparison of concordance scores for first level criteria

Criterion	Effectiveness score		Weight in sub-factor		Unweighted comparison scores		Weighted comparison scores	
	IR	ER	IR	ER	IR.ER	ER.IR	IR	ER
VI.1.A-1		12	1,00	0,50	0,50	0,33	0,50	0,17
VI.1.A-2		24	0,00	0,50	0,50	0,67	0,00	0,33
VI.1.B-1	30	12	0,38	0,43	1,00	0,00	0,38	0,00
VI.1.B-2		10	0,25	0,14	0,50	0,28	0,12	0,04
VI.1.B-3			0,38	0,43	1,00	1,00	0,38	0,43
VI.1.C-1		13	0,00	0,50	0,50	0,36	0,00	0,18
VI.1.C-2			0,00	0,50	1,00	1,00	0,00	0,50
VI.2.A-1	24	12	0,43	0,33	1,00	0,00	0,43	0,00
VI.2.A-2		8	0,14	0,33	0,50	0,22	0,07	0,07
VI.2.A-3	12	4	0,43	0,33	1,00	0,00	0,43	0,00
VI.2.B-1	24		0,43	0,20	0,67	0,50	0,29	0,10
VI.2.B-2	36		0,43	0,40	1,00	0,50	0,43	0,20
VI.2.B-3			0,14	0,40	1,00	1,00	0,14	0,40
VI.2.C-1	12		1,00	1,00	0,33	0,50	0,33	0,50
VI.3-1			0,25	0,25	1,00	1,00	0,25	0,25
VI.3-2			0,00	0,25	1,00	1,00	0,00	0,25
VI.3-3			0,50	0,25	1,00	1,00	0,50	0,25
VI.3-4	36		0,25	0,25	1,00	0,50	0,25	0,13

IR = Ireland; ER = Emilia-Romagna.

Table 9 - Hierarchical weighted sum on concordance scores

A) Aggregation by sub-Factor					B) Aggregation by factor					C) Overall score	
	Score		Weight			Score		Weight		Score	
Sub-Factor	IR	ER	IR	ER	Factor	IR	ER	IR	ER	IR	ER
Soil	0,5	0,5	1,00	1,00	Soil	0,5	0,5	0,10	0,30		
Water quality	0,9	0,5	1,00	0,54	Water	0,9	0,6	0,26	0,32		
Water quantity	0,0	0,7	0,00	0,46							
Biodiversity (flora and fauna)	0,9	0,1	0,35	0,21	Biodiversity	0,7	0,5	0,44	0,23	**0,8**	**0,6**
Biodiversity (habitat)	0,9	0,7	0,35	0,36							
Biodiversity (genetic diversity)	0,3	0,5	0,30	0,43							
Landscape	1,0	0,9	1,00	1,00	Landscape	1,0	0,9	0,20	0,15		

IR = Ireland; ER = Emilia-Romagna.

Although the final scores are not so different from the previous methodologies, the evaluation of the results across environmental factors within each case changes considerably in both areas. In particular the Soil factor in Ireland and Landscape factor in Emilia Romagna, which were set equal to zero in the previous approach, are now of some relevance in both cases (though, in both cases, they have the lowest score when compared to other factors).

Discussion

These results may be used to help inform two main themes: the information collected on the performance of AESs, and the analysis and interpretation of such information to assess AESs results. We also critique the methodology adopted here.

Based on these results, a tentative evaluation might infer that the AESs only partially achieved the local objectives and that the way in which the AESs are implemented can be reasonably improved to achieve higher effectiveness. However, the evaluation is strongly affected by the scarcity of quantitative data on actual effectiveness and any conclusions from this methodology should be quite tentative. In addition, care is required due to the lack of quantitative target levels for the objectives and the difficulty in assessing the relative importance of (numerous) different criteria. The number and variety of criteria makes it sometimes difficult to sum up the results and to come up with consistent judgements in terms of overall policy performance and trade-offs among objectives. Clearly, the ability to properly evaluate the results depends not only on the collection of a large amount of information, but also on the formalisation of a consistent evaluation framework at the design stage of the schemes.

The effectiveness criteria and MCA methodology applied in this paper, however simplified, provide insight into the difficulties and issues arising in the evaluation process. Many limitations to MCA come from the comments above. How reliable is the information contained in MTEs? Do (or can) the authors of the MTEs assess the quality of the information that they are using? This has an impact on the outcome of the MCA approach in this study, because the input information is derived from the MTEs. In addition, what are the limitations of assuming that the policy makers are correct in the targets they set and in the areas of applicability they set for measures?

The aggregation of multiple criteria and the comparison across regions is perceived as a need by EU policy makers in order to provide overall evaluations of complex schemes, in place of the wide number of criteria used for the institutional evaluation of AESs. However the choice of criteria, the possible intermediate aggregation and the mathematical complexities of the methodology might lead to a partial hiding of relevant disaggregated information for policy (though, of course, aggregation does not preclude the consideration and use of original indicator data). Partial or multiple-level aggregation can thus be suggested as ways to formulate consistent overall judgement without compromising the understanding of the drivers of such results.

Even when final aggregation and scores are achieved, a key issue is the understanding of the contribution of different variables to the overall results and interpretation. In particular, it is necessary to distinguish between the environmental variables (such as location, territorial features, etc.), the economic context in relation to local production and the institutional factors. While improved evaluation systems increase the ability to understand the results of AESs, the cost/effectiveness of such evaluation systems for policy purposes may also be considered. This suggests a need for further research steps to develop explicit consideration of further performance indexes that can aggregate environmental effectiveness and expenditure information.

Some other improvements of the present work are straightforward. First of all, the parameters adopted may be made more robust through the involvement of experts or policy makers in the evaluation, respectively, of environmental performance and weights (e.g. see Finn *et al.*, 2009). Secondly, in order to develop a full objective evaluation, environmental effectiveness should be considered in relation to the total effects potentially achievable in each area, in terms of suitability of the area to be involved in the proposed measures. Finally, further qualification of the achievements measured by the indicators should strongly consider how realistic is the match between the level of ambition of environmental targets, and the available resources.

In addition, this evaluation approach could be made more interesting if more than two case study areas were considered, with possible differentiated ranking of performance according to different criteria. This emphasises once more the need to use comparative information (with due caution), in order to put local performance in a wider perspective.

Acknowledgements

This paper presents intermediate results obtained within the EU project SSPE-CT-2003-502070 'Integrated tools to design and implement Agro Environmental Schemes' (ITAES). It does not necessary reflect the view of the European Union and in no way anticipates the Commission's future policy in this area. A previous version of this paper was presented at the XI Congress of the European Association of Agricultural Economists in Copenhagen, 23-27 August, 2005.

References

Bartolini, F. Gallerani, V. and Viaggi D. 2007. *Consolidated Report on multicriteria analysis of Agri-Environmental Schemes*. ITAES project: document n. ITAES WP10 P6 D21.

Court of Auditors. (2000). *Special Report No. 14/2000 on 'Greening the Cap' together with the Commission's replies*. Official Journal of the European Communities C 353, 43, pp. 1-56.

EEA, 2005. *Agriculture and environment in EU-15 – the IRENA indicator report*. Report No. 6/2005, Copenhagen: European Environment Agency.

European Commission, 2000. *Common evaluation questions with criteria and indicators*. Bruxelles.

European Commission, 2002a. *Common Indicators for monitoring rural development programming 2000-2006*, COM(2002) XX. Bruxelles.

European Commission, 2002b. *Guidelines for the mid term evaluation of rural development programmes 2000-2006 supported from the European Agricultural Guidance and Guarantee Fund*, Bruxelles.

European Commission, 2006. *Rural Development 2007-2013. Handbook on Common Monitoring and Evaluation Framework*, Guidance document, Bruxelles.

Finn, J.A., Bartolini, F., Bourke, D., Kurz, I. and Viaggi, D., 2009. Ex post environmental evaluation of agri-environment schemes using experts' judgements and multicriteria analysis, *Journal of Environmental Planning and Management*, 52, pp. 717-737.

Finn, J.A., Kurz, I., and Bourke, D., 2008. Multiple factors control the environmental effectiveness of agri-environmental schemes: implications for design and evaluation, Tearmann: *Irish Journal of Agri-Environmental Research*, 6, pp. 45-56.

Guitouni, A. and Martel, J. M., 1997. Tentative guidelines to help choosing an appropriate MCDA method, *European Journal of Operational Research*, 109, pp. 501-521.

Hayashi, K. 1998. Multicriteria aid for agricultural decisions using preference relations: methodology and application, *Agricultural Systems*, 58, pp. 483-503.

Hayashi, K., 2000. Multicriteria analysis for agricultural resource management: A critical survey and future perspectives, *European Journal of Operational Research*, 122, pp. 486-500.

Keeney, R. L. and Raiffa, H., 1976. *Decisions with Multiple Objectives: Performances and Value Trade-Offs,* New York: Wiley.

Maystre, L. Y., Pictet, J. and Simos, J., (eds.) 1994. *Méthodes multicritères ELECTRE*. Lausanne, Switzerland: Presses Politechniques et Universitaires Romandes.

Oréade-Brèche, 2005. *Evaluation des Mesures Agro-Environnementales*, Auzeville: France.

Primdahl, J., Vesterager, J.P., Finn, J.A., Vlahos, G. Kristensen, L. and Vejre, H. Impact models for agri-environmental schemes: examples from policy practice and potential for policy assessment and evaluation, *Journal of Environmental Management*, submitted.

Purvis, G., Louwagie, G., Northey, G., Mortimer, S., Park, J., Mauchline, A., Finn, J.A. Primdahl, J., Vejre, H., Vesterager, J.P., Knickel, K., Kasperczyk, N., Balázs, K., Vlahos, G, .Christopoulos S. and Peltola. J. 2009. Conceptual development of a harmonised method for tracking change and evaluating policy in the agri-environment: the Agri-environmental Footprint Index. *Environmental Science and Policy*, 12, pp. 321-337.

Rehman, T. and Romero, C., 1993. The application of MCA paradigm to the management of agricultural systems: some basic considerations, *Agricultural Systems*, 41, pp. 239-255.

Roy, B., (ed.) 1985. Méthodologie multicritère d'aide à la décision, Paris: Economia.

Saaty, T.L., (ed.) 1980. *The analytic hierarchy process: planning, priority setting, resource allocation*, New York: McGraw-Hill.

Zeleny, M., (ed.) 1982. *Multiple criteria decision making*, New York: McGraw-Hill.