

# PROJECT FINAL REPORT

**Grant Agreement number: 312071**

**Project acronym: ENVIEVAL**

**Project title: Development and application of new methodological frameworks for the evaluation of environmental impacts of rural development programmes in the EU**

**Funding Scheme: Collaborative project**

**Period covered: from 01/01/2013 to 12/12/2015**

**Name of the scientific representative of the project's co-ordinator<sup>1</sup>, Title and Organisation:**

**Dr. Gerald Schwarz, Johann Heinrich von Thünen Institute**

**Tel: +49 (0)531-596-5140**

**Fax: +49 (0)531-596-5199**

**E-mail: [gerald.schwarz@ti.bund.de](mailto:gerald.schwarz@ti.bund.de)**

**Project website address: [www.envieval.eu](http://www.envieval.eu)**

---

<sup>1</sup> Usually the contact person of the coordinator as specified in Art. 8.1. of the Grant Agreement.

## Table of Content

<i>1. Summary report</i>	<i>3</i>
<i>a. Executive summary</i>	<i>3</i>
<i>b. Project context and objectives</i>	<i>4</i>
<i>c. Project results</i>	<i>7</i>
<i>d. Dissemination and Impact</i>	<i>34</i>
<i>e. References</i>	<i>41</i>
<i>f. The address of the project public website</i>	<i>41</i>

## 1. Summary report

### a. Executive summary

Evaluations of environmental impacts of RDPs are characterized by a number of methodological challenges. Robust counterfactuals are critical to enable a clear attribution of observed environmental changes to implemented policy measures and programmes. This requires the consideration of sample selection issues in the design of comparison groups. However, recent methodological developments have improved the understanding and capacity of analysing the impacts of farming and forestry on the provision of public goods. Against this background, the main aim of ENVIEVAL is to test the contributions of new indicators and methods to addressing the main challenges in environmental evaluations and to develop a conceptual and methodological framework for the evaluation of environmental impacts of rural development measures and programmes in EU Member States.

The methodological frameworks and evaluation tools were tested through the application of public good case studies in selected study areas in Germany, UK, Greece, Finland, Italy, Lithuania, and Hungary. The public good case study approach (covering climate stability, biodiversity, agricultural landscapes, water quality, soil quality and animal welfare) allowed the development, testing and integration of evaluation methods according to their suitability for specific environmental objectives, and reflects the central aim of the Common Agricultural Policy (CAP) to deliver public goods from farming and forestry.

ENVIEVAL included an economic cross-cutting component assessing the cost-effectiveness of the evaluation methods developed and tested for use in the methodological framework. The cost of developing and applying the different indicators, monitoring requirements and evaluation methods and their impacts on the quality of the evaluation results were compared and tested in the public good case studies, considering the robustness of the results, the level of details and the ability to draw generic conclusions. This project component provided valuable information for evaluators and policy-makers on the suitability and selection of different evaluation methods in future evaluations taking into account differences in data availability between Member States and environmental aspects, skills of the evaluators and existing financial resources.

The policy and end-user relevance of the outputs of ENVIEVAL were delivered through workshops and close collaboration with European, national and regional evaluators, managing authorities and other stakeholders working with the partners. A separate work package was allocated to stakeholder involvement and dissemination. A Dissemination and Impact Committee coordinated and oversaw this process. The successful collaboration with the European Evaluation Helpdesk, national and regional evaluators, managing authorities and other relevant stakeholders in previous rural development projects facilitated this important process. A user-friendly, methodological handbook synthesises fact sheets on the development and application of the different evaluation tools and provides guidance to evaluators and policy-makers for evaluations of EU rural development programmes in the enhanced AIRs in 2017 and 2019 as well as the ex-post evaluation in 2024.

## ***b. Project context and objectives***

ENVIEVAL developed and tested improved tools for the evaluation of environmental impacts of rural development measures and programmes in EU Member States. In order to achieve this main aim, the project had five objectives:

- To review implemented rural development programmes, existing monitoring and indicator systems, and new methodological developments in environmental policy evaluation
- To develop new methodological frameworks for the evaluation of net environmental effects of rural development programmes against their counterfactual
- To test and validate the selected evaluation methods through public good case study applications in the partner countries and close collaboration with the European Evaluation Network, national and regional evaluators and managing authorities
- To assess the cost-effectiveness of the tested indicators and evaluation methods
- To provide a methodological handbook for the evaluation of environmental impacts of rural development programmes.

The monitoring and evaluation system for Rural Development Programmes (RDPs) for the period 2014 – 2020 is set out at different levels by the Regulation (EU) No 1303/2013, Regulation (EU) No 1306/2013 and Regulation (EU) No 1305/2013. For the first time, the current programming period (2014-2020) offers a Common Monitoring and Evaluation Framework (CMEF), providing rules and procedures necessary for evaluating the whole CAP, and a Common Monitoring and Evaluation System (CMES), defining the rules and procedures within the CMEF which specifically relate to rural development (Pillar II of the CAP). Quantitative common indicators applicable to each programme are defined. Since common indicators may not fully capture all effects of programme activity, for example in relation to national priorities and site-specific measures, it is necessary that Member States and programme partnerships define additional and programme specific indicators for each type of indicator in a flexible manner, but in accordance with the general principles of the CMEF and CMES. Evaluations of results of RDPs have to be reported for the first in the extended Annual Implementation Report (AIR) in 2017, followed by impact evaluations in 2019 and the ex-post in 2024 (European Evaluation Helpdesk, 2015). This approach is more strategic and consistent than earlier evaluation approaches. However, significant issues remain: (i) the linkages between the different levels of indicators (e.g. from result indicators at measure and axis level to impact indicators at programme level); (ii) the linkages between indicators and different rural development measures; (iii) the complexity and data requirements of existing and additional impact indicators; (iv) counterfactual development for measures implemented across large areas; and (v) the quantification of net impacts of the Programmes at the macro level and establishing causal-effects relationships.

Environmental impacts of rural development measures are strongly influenced by site-specific circumstances, may take a long time to emerge and often depend on a range of other intervening

factors. Recent methodological developments, for example in environmental farm planning and modelling, environmental impact assessments, life cycle assessments, spatial econometrics, regional modelling of farming, and mixed method case studies, have improved the understanding and capacity of analysing the impacts of farming and forestry on the provision of public goods in different rural environments. While it is widely recognised that the evaluation of environmental impacts of RDPs still has a number of problems, including the assessment of their environmental performance (e.g. Viaggi et al., 2015), recent advances in the development of indicators, data availability and geographic analysis provide new opportunities to address existing key challenges of the CMEF.

The ENVIEVAL project integrated improved and advanced evaluation methods into new methodological frameworks to evaluate environmental impacts of rural development programmes at micro- and macro-levels. The main innovative aspects of the new logic model-based methodological frameworks are that they enable the integration of micro- and macro-level evaluations (and their results) and provide guidance on the selection and application of cost-effective evaluation methods to estimate net effects of rural development programmes on the different main public goods from farming and forestry. The integration takes into account the intervention logic of the CMEF and CMES, data and monitoring requirements of a coherent set of indicators, and their cost-effectiveness in estimating net effects in relation to specific environmental objectives and public goods.

ENVIEVAL covered a set of EU Member States of Germany, UK, Greece, Finland, Italy, Lithuania, and Hungary, within which it chose regional study areas to test the suitability of methods to evaluate the impacts of the rural development programmes on different environmental public goods. The partner countries were drawn from across the four main classes of the Köppen–Geiger climate classification system (Peel et al., 2007) of: dry-summer subtropical or Mediterranean climates; humid continental climate (warm summers and cold winters); temperate oceanic climate; and, cool continental (or subarctic) climate. The partner countries cover a wide range of different environmental, socio-economic and political characteristics of rural areas. The state and extent of the provision of different public goods from agriculture such as biodiversity, water quality and landscapes vary greatly across the different rural environments in the partner countries, as do the priorities in the rural development programmes, so providing a menu of different key rural development measures across all axes. Agricultural systems vary from intensive farming with fertile soils and favourable climatic conditions (e.g. parts of Germany, Italy and the UK), to extensive livestock systems in some of the most marginal and remote areas in the EU which also suffer from unfavourable natural conditions and isolation from markets (e.g. remoter areas of Finland and Greece). Agricultural sectors in the Baltic States and Hungary (new Member States) are going through a process of significant structural change affecting the quality and quantity of public goods they provide. The differences in the provision of public goods, rural development programmes and agricultural structures provide a diverse setting for the development and testing of new and improved tools to evaluate the environmental impacts of rural development programmes in a set of case studies in the partner countries which will also take account of different data requirements and availability.

The methodological frameworks and evaluation tools were tested through the application of public good case studies in selected study areas in the partner countries. The public good case study approach allowed the development, testing and integration of evaluation methods according to their suitability for specific environmental objectives, and reflects the central aim of the Common Agricultural Policy (CAP) to deliver public goods from farming and forestry. The selection of the public good case studies built on their relevance to farming and forestry, as identified in European Network for Rural Development (ENRD) (2011) and Institute for European Environmental Policy (IEEP) (2010), with respect to the environmental objectives of CAP and the structure of the CMEF.

In addition to the environmental public goods of climate stability, biodiversity, agricultural landscapes, water quality and soil quality, the project paid particular attention to animal welfare and included an animal welfare case study. Animal welfare is one of the key objectives of the rural development regulation (European Commission, 2005) and the Treaty of Lisbon (Article 13; European Commission, 2009), but is currently not explicitly covered in the CMEF. Some Member States have included additional impact indicators, but one of the main difficulties in the evaluation of impacts on animal welfare of rural development programmes is the formulation of robust and quantifiable indicators for different livestock species (and a lack of underlying data) suitable for policy evaluation. ENVIEVAL has reviewed animal welfare indicators and qualitative evaluation methods in public good case studies, which built on recent and current developments in a number of international animal welfare projects (e.g. Welfare Quality, AWARE and AWIN projects).

ENVIEVAL included an economic cross-cutting component assessing the cost-effectiveness of the evaluation methods developed and tested for use in the methodological framework. The cost of developing and applying the different indicators, monitoring requirements and evaluation methods and their impacts on the quality of the evaluation results were compared and tested in the public good case studies, considering the robustness of the results, the level of details and the ability to draw generic conclusions. This project component provided valuable information for evaluators and policy-makers on the suitability and selection of different evaluation methods in future evaluations taking into account differences in data availability between Member States and environmental aspects, skills of the evaluators and existing financial resources.

The policy and end-user relevance of the outputs of ENVIEVAL were delivered through workshops and close collaboration with EU level stakeholders, evaluators, managing authorities and other stakeholders working with the partners. A separate work package was allocated to stakeholder involvement and dissemination. A Dissemination and Impact Committee coordinated and oversaw this process. The successful collaboration with the European Evaluation Helpdesk, evaluators, managing authorities and other relevant stakeholders in previous rural development projects facilitated this important process. A user-friendly, methodological handbook synthesises fact sheets on the development and application of the different evaluation tools and provides guidance to evaluators and policy-makers for future evaluations of EU rural development programmes.

### *c. Project results*

#### *The ENVIEVAL approach*

The state and extent of the provision of different public goods from agriculture such as biodiversity, water quality, landscapes and animal welfare, as well as the priorities in the rural development programmes, vary greatly across the different rural environments in the partner countries including Finland, Germany, Greece, Hungary, Italy, Lithuania and UK. The differences in the provision of public goods, rural development programmes and agricultural structures provide a diverse setting for the testing of improved tools to evaluate the environmental impacts of rural development programmes in a set of case studies which will also take account of different data requirements and availability.

Figure 1 outlines the integration of different tasks which were required to develop and test the methodological framework. In a first step, suitable indicators and recent methodological developments for counterfactual evaluation of environmental impacts at micro and macro level were identified (WP2 – WP5) and their potential to address future evaluation challenges and needs were discussed with relevant stakeholders (WP9). Data requirements of the selected indicators and methods were assessed and case study areas with good data availability selected. The selection built on the availability of the data required to test the different indicators and methods and on their relevance to farming and forestry, with respect to environmental objectives of CAP and the structure of the CMEF and CMES.

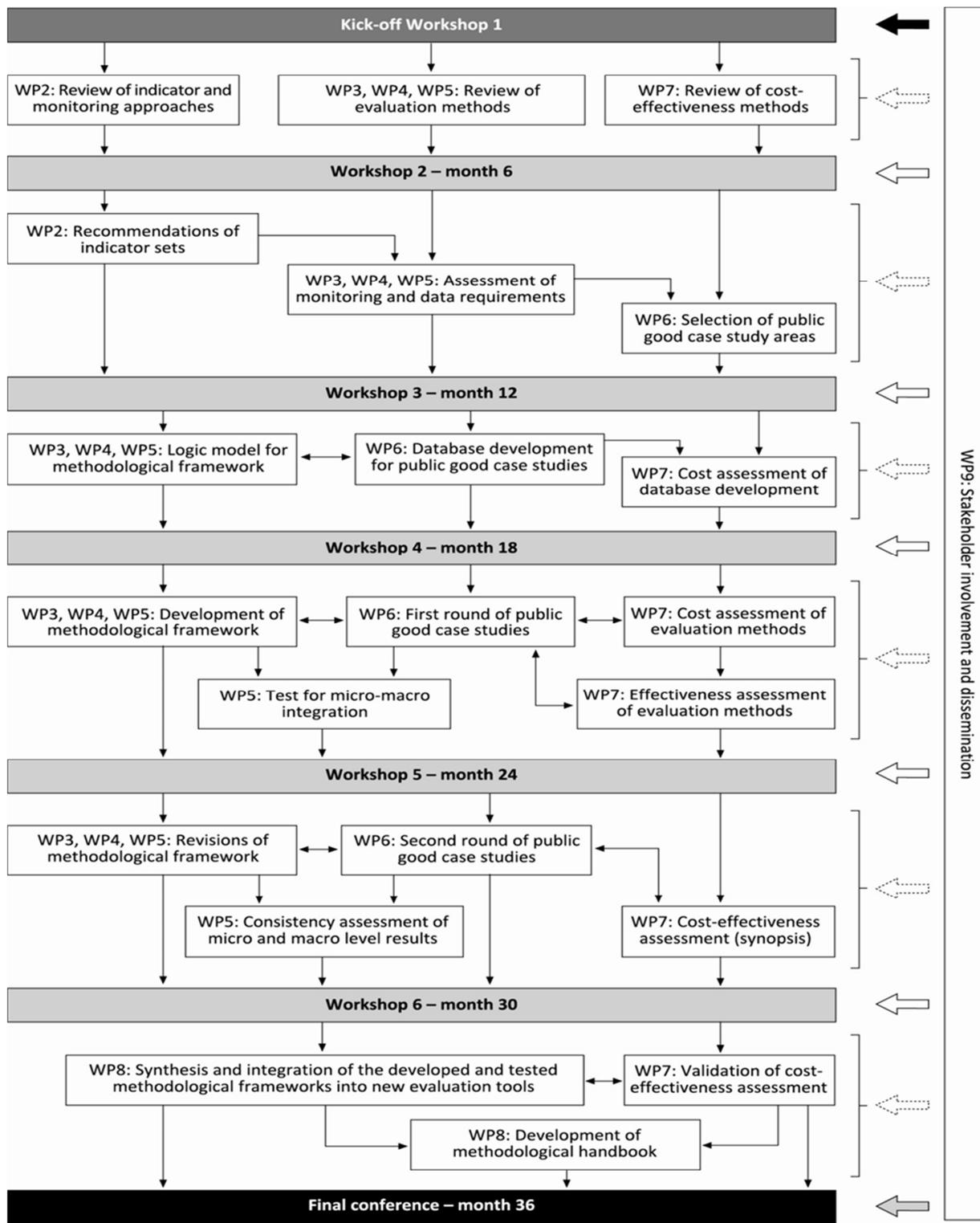
An important conceptual step was then the development of logic models for the methodological framework for the evaluation of net environmental effects of rural development programmes against their counterfactual in WP3 – WP5. The logic model provided the conceptual framework and a decision tree for evaluators and managing authorities to develop a consistent methodological framework combining the selection of indicators, counterfactual approaches and micro and macro-level evaluation methods in accordance with the specific circumstances facing the evaluator or managing authority. The logic models form the basis for the methodological framework developed. The practical relevance of the logic models, as well as the case study design, was reviewed and validated in national and international stakeholder consultations. This included different sets of interviews and workshops, organised in collaboration with the European Evaluation Helpdesk.

Based on the results of case study testing in WP6, the methodological framework was revised and validated with stakeholders. The final methodological frameworks include a recommended set of methods for use in assessing net effects at micro and macro levels based on consistent counterfactuals. The results of the case study testing provided also the basis for the indicator and method fact sheets in the handbook as well as the lessons learnt for the policy briefs.

A particular aspect of ENVIEVAL was the assessment of the cost-effectiveness of the evaluation approaches in WP7. Guidelines were produced to collate the required data from the case studies in

order to identify and quantify the different cost components, classify costs of the developed evaluation in absolute and relative terms, and analyse the main determinants of the costs of the tested evaluation methods. The effectiveness was assessed in terms of their impacts on the quality of the evaluation, using a set of standardised judgement criteria. The categorisation was carried out with stakeholders using participatory approaches in workshops in each partner country. In the last step a cost-impact synopsis was carried out to compare and assess the cost-effectiveness of the tested indicators, monitoring requirements and evaluation methods and to inform the fact sheets for the methodological handbook in WP8.

In the project synthesis, fact sheets of the tested indicator and methods and policy briefs were produced and reviewed by the stakeholders before and during the final project conference. Draft concepts of the methodological handbook were discussed with the stakeholders and the assessments of the different methodological approaches for the key steps of the environmental evaluation process (e.g. indicator selection, counterfactuals, micro level, macro level) formed the basis for the methodological handbook for the evaluation of environmental impacts of RDPs. The handbook provides step-by-step guidance on designing cost-effective evaluation approaches taking in the experiences from the case study testing and the feedback from the different stakeholder consultations. The handbook includes the fact sheets which synthesise key aspects of the tested indicators and methods such as description of the indicator / method, data and skill requirements, consideration of counterfactuals, context of the case-study testing, strengths and weaknesses of the indicator / method and their recommended application. The main results of the different parts of the ENVIEVAL project are synthesised in the remaining part of section 1c.



- ← Kick-off workshop (Partners and Stakeholder Reference Group)
- ← Workshops with partners and Stakeholder Reference Group
- ← Mobilisation of specific stakeholders through technical meetings and questionnaires
- ← Final conference to a broader stakeholder group

Figure 1 Overview of the ENVIEVAL approach

### ***Indicator and method reviews to inform the development of the methodological framework***

Various indicator and monitoring frameworks have been proposed but also used for more than 20 years. In order to review these attempts, an inventory of indicators was created including the ones used in 16 member states' evaluation documents examined for the project by all partners and subcontractors, based on a common reporting template. The objective of the review in WP2 was to recommend suitable evaluation indicators to be incorporated into the methodological frameworks of the evaluation tools and tested in the public good case study areas.

The key finding of the summary report reveals the absence of indicators to assess the impact between certain combination of measures and public goods, especially in Axes 3 'Quality of life and Diversification' and 4 'Leader', even though an influence has been specified. Therefore for the suggestion of suitable indicators priority needs to be given to those cases, where particular gaps were identified.

The process of selecting indicators for RD measures that lack indicators can be broken down into the following stages:

- a) Identification of combinations where there was a gap.
- b) Using the lists of indicators for each public good that were built on the review of evaluation reports (see Appendix A of Deliverable D2.1 on [www.envieval.eu](http://www.envieval.eu)), the most relevant indicators were suggested. The criteria of suggestion have been based on the relation of the general objectives of the measures as well as the measurement of the indicators and their characteristics.
- c) In some cases there was need for adaptation of the proposed indicator to the specific measure. In those cases a comment indicates this.
- d) Several indicators have been estimated using different approaches by different authorities. In those cases all alternative approaches for the estimation are presented.

The outcome of this approach can address the issues concerning the absence of indicators and was summarised in tables linking RD measures and proposed indicators for each public good. Moreover, since the common indicators defined by the CMES are not always detailed or specific enough to reflect the wider benefits of a measure, there is a need for additional and more flexible indicators dealing with site-specific circumstances. In order to exploit the potential offered by other indicator frameworks suggested in studies or research projects as well as the latest version of context, result and impact indicators provided by the Commission services, an effort was made to examine them and construct a list of alternative suitable indicators per public good. Indicators in this list are classified into sub-categories according to their relevant farming/environmentally features (published in Appendix C of deliverable D2.1). The final selection of appropriate indicators was done in the context of the selected evaluation method, data availability and environmental circumstances in each case study area.

In parallel to the indicator reviews, reviews of methodologies for counterfactual development and the assessment of environmental impacts of rural development programmes and measures at micro and macro levels have been carried out in WP3 – WP5 with the aim to summarise strengths and weaknesses of candidate methods for the case study testing and the methodological framework developed in ENVIEVAL.

The use of well-established quantitative methods such as the Propensity Score Matching (PSM) and Propensity Score Matching Difference-in-Difference (PSM-DD) is recommended for counterfactual-based environmental impact assessment (WP3). Both of these methods can overcome the biases suffered by naïve estimators given that sufficient information exists on the control group (i.e. non-participants) and time-related factors. It is important that baseline scenarios are created before the implementation of a programme to ensure an easier way to construct a counterfactual at the evaluation phase.

At micro level (WP4), five main categories of methodologies/models were identified including sustainability indicators, statistical sampling and monitoring, bio-physical and spatial models, agent-based modelling and integrated models (with modules at farm level).

These categories are not mutually exclusive and present some overlaps. For example, integrated models are based on empirical evidence coming from statistical sampling and monitoring; sustainability indicators could result from the outputs of integrated models; while biophysical models and spatial analysis are frequently considered within integrated models. Based on the literature review, the strengths and weaknesses of these methodologies in terms of RDP environmental impacts assessment were assessed. Each method has been scored according to their relevance in terms of micro-level applicability, based on four strengths and three weaknesses that are common to each selected methodology. Furthermore, relevant specific strengths and/or weaknesses have been added to describe comprehensively the suitability at micro level.

Table 1 Categories of methods at farm (micro) level – Strengths and Weaknesses

Categories of models	Strengths	Relevance	Weaknesses	Relevance
<b>Sustainability Indicators</b>	Micro/macro linkage	++	Limited availability of farm level data	++
	Applicability to counterfactual analysis	++	Lack of consistency, robustness and representativeness of micro level data	++
	Bottom up approach	+++	Not applicable to all public goods	++
	Interaction and integration with other models	+++	More oriented to pressure indicators (farm practices instead of actual environmental impacts)	+++
	Possibility to obtain composite indicators	+++		
<b>Ecological</b>	Micro/macro linkage	+	Limited availability of farm level	+

<b>Footprint</b>			data	
	Applicability to counterfactual analysis	++	Lack of consistency, robustness and representativeness of micro level data	+
	Bottom up approach	+	Not applicable to all public goods	+
	Interaction and integration with other models	+	Complexity models require high-skilled evaluators	++
	Measure of the environmental impacts of the Life Cycle (e.g., energy, water)	+++	Strict requirements related to data processing	+++
			Strictly dependent by the selected productive system	+++
<b>Statistical Sampling</b>	Micro/macro linkage	++	Limited availability of farm level data	++
	Applicability to counterfactual analysis	+++	Lack of consistency, robustness and representativeness of micro level data	+++
	Bottom up approach	++	Not applicable to all public goods	+++

At macro level (WP5) the most important methodological developments are the advances made in relation to multi-criteria, spatial analytical approaches and integrated approaches, as well as efforts made to address the scale mismatch between economic and ecological/natural sciences. These developments are able to contribute to addressing the challenges posed by the demand for measuring the impact of RDP activities/investment on the delivery of public goods. Methods reviewed include statistical, hierarchical, spatial analytical, multi-criteria and integrated methods. Table 2 below provides an overview of their strengths and potential contributions to the main evaluation challenges.

Both the review of the evaluation reports and the interviews with the evaluators showed that complex methods and models have rarely been used in past evaluations. Hence, a potential lack of experience and methodological skills in using complex quantitative methods for environmental evaluations had to be considered in the selection of case study methods and the development of the methodological framework. The importance of different stakeholder aspirations and capacities across the EU for the comprehensiveness and quality of RDP evaluations was also raised during the stakeholder workshop. The suitability of the selected candidate methods for case study testing, and consequently for inclusion in the methodological framework, was considered under different circumstances with respect to data availability, and stakeholder aspirations and capacities in the different member states.

Table 2 Overview of reviewed methods at macro level and their key strengths

Method	Key methodological aspect	Addressed challenge(s) or improvement of evaluation
<b>Statistical</b>	Multi-scale/dimension trend analysis	determination of true causation
	GLM Modelling	determination of true causation
	Scaling	data/model integration for aggregation RDP impacts
	Down scaling using generalised additive modelling (GAM)	data integration for aggregation RDP impacts
	Up scaling	data integration for aggregation RDP impacts
<b>Hierarchical</b>	Hierarchical sampling	assessment of Impact on the wider context (public goods) – true causation
	Cluster analysis, regression	assessment of Impact on the wider context (public goods) – true causation
	Statistical modelling at different spatial resolutions	assessment of Impact on the wider context (public goods) – true causation
<b>Spatial analytical</b>	Spatial econometrics	assessment of net impact on the wider context (public goods) – true causation
	Landscape analysis Landscape metrics and zoning	assessment of Impact on the wider context (public goods) – true causation
<b>Multi-criteria</b>	Survey/panel Knowledge elicitation	aggregation of all economic, social & environmental effects of RDP
	Multi-Criteria Analysis	aggregation of all economic, social & environmental effects of RDP
	Life Cycle Analysis	aggregation of all economic, social & environmental effects of RDP
	Ecological footprint analysis	aggregation of all economic, social & environmental effects of RDP
<b>Integrated</b>	Expert assessment/ Knowledge elicitation	aggregation of all economic, social & environmental effects of RDP
	Mixed case studies	aggregation of all economic, social & environmental effects of RDP
	Model integration	aggregation of all economic, social & environmental effects of RDP
	Mapping landscape functions and services	Quantification of change in the value of a public good
	Adaptive Management	Quantification of change in the value of a public good per natural region/landscape character

### ***Testing the selected indicators and methods to improve environmental evaluations of RDPs***

The case studies in WP6 are the central tool to validate the developed methodological framework for the counterfactual-based evaluation of environmental impacts of RDPs at micro and macro level and to test the contributions of indicators and methods identified in previous reviews and theoretical analyses to address the main challenges in evaluations of environmental impacts of RDPs. Of the main environmental public goods identified in ENRD (2011), the case studies focus on climate stability, biodiversity, water quality, soil functionality and cultural landscapes. These environmental public goods reflect the key environmental objectives of the CAP and are at the core of the needs of evaluations of environmental impacts of the rural development programmes in the Member States. In addition to the testing of indicators and methods in the context of environmental public goods, a review of the integration of animal-based (result-based) indicators into a multi-criteria evaluation framework of animal welfare has been carried out in a final case study deriving guidelines for the selection of animal welfare indicators.

The selection of case study areas is provided in Deliverable D6.1 published on [www.envieval.eu](http://www.envieval.eu). Table 3 provides an overview of the public good case studies summarising the main evaluation challenges addressed in the case studies, the case study context, the indicators and methods tested and their expected outcome. The contributions of the case study testing of the selected indicators and methods could be classified into three main contributions:

- Contributions of additional (non-CMES) indicators tested to address indicator gaps
- Contributions of advanced modelling approaches tested at micro and macro level for dealing with the complexity of public goods, considering other intervening factors and providing solutions for situations without (or very limited) non-participants
- Contributions to the integration of counterfactuals and sample selection issues in environmental evaluations of RDPs.

Table 3 Overview of the public good case studies

Case study	Public good	Main evaluation challenges addressed	Case study context		Indicators tested		Methods tested		Expected outcome
			Case study area	Policy measures	Micro level	Macro level	Micro level	Macro level	
BW-HU	Biodiversity wildlife	Establishing robust causal linkages between implemented policy measures and changes in biodiversity indicators at micro level	Heves-plain	214	Number of farmland bird species (NBS) and Number of farmland bird individuals	Farmland Bird Index	Spatial analyses of survey spots (with and without comparison groups)	Spatial analyses of quadrats (with and without comparison groups)	<ul style="list-style-type: none"> <li>• Assessment of additional biodiversity indicator (NBS) at micro level</li> <li>• Set up feasible hierarchical sampling strategies which allow consistent biodiversity impact assessment at micro and macro levels based on accessible data</li> </ul>
BW-LT	Biodiversity wildlife	Indicator gaps at micro level - missing robust indicators to evaluate net-effects of specific agri-environmental measures on biodiversity wildlife at micro level; Lack of coordination and integration of available data between agriculture and environmental sectors	Micro level assessment: Šilutė and Vilnius regions	214	Singing males of corncrake (corncrake density) , white stork breeding density and breeding success		Multiple regression analysis	Hierarchical sampling and spatially explicit up-scaling	<ul style="list-style-type: none"> <li>• Assessment of additional biodiversity indicator at micro level</li> <li>• Assessment of appropriateness of biodiversity monitoring data for the AEM evaluation</li> <li>• Recommendations for improving integration of data systems at public administration in order to enable cross-sectoral linkages and impact assessments of causal linkages between agricultural and environmental sectors.</li> </ul>
CC-FI	Climate stability	Area-wide implementation of policy measures and lack of a non-participant control group.	Finland: whole country	211, 212 and 214	CO <sub>2</sub> equivalent measures both with and without land cover changes		Not applicable	Partial equilibrium model	<ul style="list-style-type: none"> <li>• Application for macro level evaluation in particular in cases with limited or lacking non-participants</li> <li>• Regionally differentiated sectoral modelling framework to consider indirect effects at macro level</li> </ul>
CC-IT	Climate stability	Complexity and data requirements of existing and additional impact indicators and public good assessments	Emilia Romagna	214, 216, 221	GHG balance at farm level/at crop level		IPCC and LCA approaches (Carbon footprint)	Up-scaling from hierarchical sampling (consistency check)	<ul style="list-style-type: none"> <li>• Assessment of the suitability and robustness of the footprint method to evaluate net-effects of RDPs</li> <li>• To infer regional result (macro level) to evaluate RDP environmental impact in terms of carbon emissions</li> <li>• Assessment of carbon emissions (CO<sub>2</sub>) in different agricultural contexts and processes (farm type)</li> </ul>

Case study	Public good	Main evaluation challenges addressed	Case study context		Indicators tested		Methods tested		Expected outcome
			Case study area	Policy measures	Micro level	Macro level	Micro level	Macro level	
HNV-IT	High Nature Value farmland	Establishing consistent micro-macro linkages to inform the net-impact assessment at micro and macro level Lack of control groups to establish advanced and robust counterfactuals to assess the environmental impact of RD measures.	Veneto Region	211, 214, 216, 221	% of HNV farmland, HNV score		Multicriteria analysis	Up-scaling of micro level results (consistency check)	<ul style="list-style-type: none"> <li>• Quantification of HNV farmland under different data availability</li> <li>• Quantitative assessment of the contribution of RDP measures to improve the diffusion of HNV farmland</li> </ul>
HNV-LT	High Nature Value farmland	Indicator gaps - Testing additional result indicators for HNV assessments covering forestry	Lithuania: Panevezys plain area	214, 221, 223	Landscape heterogeneity (changes in diversity of ecotones)		Spatial statistics	Spatial statistics	<ul style="list-style-type: none"> <li>• Quantitative assessment of the contribution of RDP measures to improve the diffusion of HNV farmland and forestry.</li> <li>• Assessment of additional HNV indicator to improve integration of forestry</li> </ul>
L-GR	Landscape	Indicator gaps - lack of suitable result and impact indicators for counterfactual-based micro and macro level assessment Establishing robust causal linkages between implemented policy measures and landscape changes in a specific traditional and local context	Island of Santorini	214 and the special measures in favour of the Small Aegean Islands	Land cover change (conversion at different points in time), Visual amenity		Spatial analysis (DiD)	Spatial analysis upscaled	<ul style="list-style-type: none"> <li>• Assessment of suitability and robustness of additional result indicators for landscape changes in a specific local context</li> <li>• Assessment of synergies of two different policy measures, including a multiple objective RD measure, not directly focusing on landscape protection but rather on maintaining the rural society.</li> </ul>
L-SCO	Landscape	Indicator gaps - lack of suitable result and impact indicators for counterfactual-based micro and macro level assessment Consistent micro-macro linkages to contribute to net-impact assessment.	Aberdeen-shire	212, 214	Coherence (shape and edge metrics), disturbance (Patch metrics) visual scale (openness/closed-ness)	Complexity (Shannon index), historicity (presence of historic features), naturalness (% UAA Natura2000)	Landscape metrics (patch/field metrics for holdings), spatial analysis (visibility)	Spatial analysis and landscape metrics (shape metrics)	<ul style="list-style-type: none"> <li>• Identification of LCA indicators that can measure impact of RDP on landscape at micro and macro level,</li> <li>• Robustness of landscape metrics, in particular the indicators for landscape character assessment, under different data availabilities</li> <li>• Systematic approach to identify causal links between implementation at micro level and impact at macro level.</li> </ul>

Case study	Public good	Main evaluation challenges addressed	Case study context		Indicators tested		Methods tested		Expected outcome
			Case study area	Policy measures	Micro level	Macro level	Micro level	Macro level	
SQ-HU	Soil quality	Explicit consideration of other intervening factors and the establishment of robust causal relationships Consistent micro-macro linkages to contribute to net-impact assessment	Hungary: whole country	214	Soil quality: Soil organic carbon (SOC) content; decrease in soil organic matter (SOM) Soil erosion: annual average soil loss per ha	Biophysical modelling approach: CLUE & USLE prediction model	Biophysical model results scientific sampling and spatially explicit up-scaling	<ul style="list-style-type: none"> <li>• Modelling based consideration of other intervening factors at macro level.</li> <li>• Comparative assessment of model-based and sampling-based approaches</li> <li>• Establishment micro-macro linkages through aggregation of spatially explicit GIS data</li> </ul>	
SQ-SCO	Soil quality	Consideration of other intervening factors and local environmental characteristics Establishment of robust causal relationships	Grampian region	212, 214, 221, 223, 225	Soil organic carbon content; soil carbon in arable land; soil erosion: annual average soil loss and sediment retention	USLE prediction model and InVest model	USLE model: calculation of the soil indicators	<ul style="list-style-type: none"> <li>• Robustness of the tested indicators and methods with limited data availability</li> <li>• Modelling based consideration of other intervening factors at macro level.</li> <li>• Establishment of robust micro-macro linkages through aggregation of spatially explicit GIS data</li> </ul>	
WQ-DE	Water quality	Lack of control groups to establish advanced and robust counterfactuals	Lower Saxony	214 and 114	GNB and Nmin	Biophysical models, PSM and regression analysis	biophysical model results, spatially explicit up-scaling	<ul style="list-style-type: none"> <li>• Testing of an alternative impact indicator (non-CMEF)</li> <li>• Integration of data from different sources to construct robust counterfactuals and use advanced statistical methods to assess net-effects</li> </ul>	
WQ-FI	Water quality	Area-wide implementation of policy measures and lack of a non-participant control group.	Southern Finland	214	Nitrogen reduction (calculated GNB, nitrogen)	Biophysical and structural modelling	Biophysical and structural model results up-scaled	<ul style="list-style-type: none"> <li>• Existing structural models provide solutions for counterfactual analysis without comparison groups.</li> <li>• Needs and capabilities to accommodate structural models with new data</li> </ul>	
WQ-GR	Water quality	Establishment of robust causal relationships and counterfactuals Lack of sufficient data measured at the required spatial level with appropriate frequency in time	Karditsa Plain of Thessaly	214	GNB & Water use/ha	Biophysical model	Biophysical model up-scaled	<ul style="list-style-type: none"> <li>• Assessment of causal relationships and solutions for counterfactuals with limited data availability</li> <li>• The case study testing is expected to address the lack of sufficient data measured at the required spatial level with appropriate frequency in time.</li> </ul>	
AW-DE	Animal welfare	Indicator gaps - lack of robust and quantifiable animal welfare indicators in RDP evaluations	North-Rhine Westphalia	215	Management and animal based indicator combination	Not applicable		Development of guidelines for the selection of robust animal welfare indicator combinations	

### *Contribution of tested additional (non-CMES) indicators*

The CMES does not provide common impact indicators for the landscape and animal welfare public goods. Evaluators and managing authorities are also required to define additional environmental result indicators to bridge the gap between evaluating effects at focus-area level and the use of impact indicators at programme level. In particular, the case studies for the public goods biodiversity wildlife, HNV and landscape, and animal welfare focussed on testing alternative and additional result and impact indicators, based on the findings of the indicator review in Deliverable D2.1. Additional indicators were also explored for water quality.

The specific biodiversity wildlife indicators - corncrake density and white stork breeding success - have been tested in Lithuania as additional result indicators being applied in addition to the FBI. Corncrake density is a suitable indicator for the evaluation of specific grassland-related agri-environmental measures at a local level and has a good responsiveness to management changes of grassland habitats. The results of the case study in Lithuania indicate that the indicator of white stork breeding success can be applied at regional and national levels for a wider range of measures. Spatial aspects of the indicator species and the use of existing monitoring programmes are key factors determining the counterfactual assessment of the effects of relevant measures under the focus area 4a Biodiversity. The example of the white stork also highlights that the consideration of socio-cultural aspects (positive image of the species and official national species of Lithuania) in the selection of the indicator facilitates good acceptance amongst farmers and other stakeholders, and consequently the availability of monitoring data through volunteers and farmers.

In the Hungarian biodiversity wildlife case study, the indicators of the number of farmland bird species and number of farmland bird individuals were developed for assessing the micro-level effects of measures under focus area 4a. The indicators are more sensitive to micro-level effects than the FBI, as the unit of analysis is linked to distinct parcels of contracted or not contracted areas. The results of the case study indicate a good responsiveness to land-management changes defined in the prescriptions of relevant measures. However, the assessment of net effects is more data intensive than for the other two indicators, and requires substantial monitoring data with survey points at a suitable spatial distribution for participants and non-participants. Overall, the indicator has good potential to be applied in other member states and programme areas where sufficient baseline data of the FBI indicator are available.

The landscape case studies tested a range of different indicators for the counterfactual assessment of the effects of relevant measures under the focus area 4a including Landscape structural and visibility indicators and the method Landscape metrics (in Scotland) and Land cover change and Visual amenity (in Greece). A land cover change indicator based on Google Earth images was tested, which provided a reasonable database for detecting landscape change. However, specific landscape features such as terraces and boundary walls are not represented and it requires a ground-level familiarity with the study area to assess changes in these features. The indicator needs to be adapted to relevant

land cover for the specific evaluation case by constructing a site-specific land-cover classification. Adoption of other commonly-used land-cover and/or landscape classifications (CORINE, EEA) might not be possible for smaller areas with rather specific landscape elements, such as traditional vineyards in Greece or traditional olive groves in Spain.

Many scientific studies have explored the assessment of the visual quality of landscapes. In the case-study testing, the indicator visual amenity also had to be adapted by the team to reflect the particular visual features of the specific landscape of the case study area. The adaptation consisted of the arbitrary assignment of values to land-cover types, which entails the risk of non-comparability across different applications.

The use of data from IACS for uptake of RDP measures, and spatial analysis of their content and change enabled a multi-dimensional assessment of impacts on the character of landscapes in the case study in Scotland. The approach used a theoretically-grounded approach which relates to landscape concepts and character and enabled causal relationships to be identified. Changes in the visibility of land cover and uses associated with selected measures, in the context of landscape character, enables temporal assessments. The use of changes in the landscape spatial metrics of land cover and use associated with RDP measures provides a second dimension for interpretation with respect to landscape character. Combinations of the three approaches enable the assessment of a broader set of net effects and better capture the complexity of environmental relationships with respect to the character of the landscape and thus the public good.

The German water quality case study explored the application of the indicator Mineral N content in the soil in autumn (Nmin). The Nmin indicator is based on well-documented, theoretically-sound models and methods. The autumn Nmin values have a strong relation to the potential nitrate that is leached into the groundwater in winter. The indicator and its characteristics are well known and used for monitoring purposes related to drinking water protection by the managing authorities. The indicator can be used as a result indicator contributing to statistical evidence of the effects of rural development measures under focus area 4b (water management) on water pollution by agricultural land use. The suitability of the indicator for statistics-based approaches (e.g. such as propensity score matching) to consider sample selection issues depends on the availability of, and access to, sufficient annual monitoring data. It is recommended to use the indicator in combination with the CMES impact indicator GNB which is well-known and widely used for monitoring water quality.

The animal welfare case study focussed on the review of suitable animal welfare indicators. The CMES does not provide guidance on animal welfare indicators. The evaluation of animal welfare impacts under the focus area 3a requires appropriate concepts to cover different animal welfare criteria targeted by relevant policy measures such as animal welfare payments and farm investment support. The case study tested the integration of a result-based approach with animal-based indicators into the evaluation of animal welfare impacts. The integration of specific animal-based indicators provides a practical solution to add a direct assessment of health criteria to the assessment

of housing and feeding criteria through the use of resource or management-based indicators. Indicators such as lameness and body conditions have a high acceptability of both stakeholders (including farmers and monitoring organisations and managing authorities) and scientists. Practitioners and farmers viewed had concerns about the use of the indicator mortality rates, as they felt that on small farms the occurrence of one accident or disease could already affect their eligibility for payment. This problem can however be solved by using average mortality rates over several (e.g. three) years.

The results of the case study indicate robust causal relationships between policy measures and animal-based indicators. Application of the indicators is recommended in a multi-criteria assessment in combination with resource- and management-based indicators. The cost-effective application depends on available monitoring data in livestock databases such as the HIT database in Germany. Few cases exist where livestock monitoring data are collected as part of animal welfare payments. High monitoring requirements and costs might prohibit the application if no data sources exist.

*Contributions of tested advanced modelling approaches at micro and macro level: Complexity of public goods and consideration of other intervening factors*

A number of advanced modelling approaches were tested for the suitability to contribute to net impact assessment at micro and macro level. Generally, the case studies tested environmental modelling approaches which require a combination with statistical methods to assess the net effects of RDP measures and approaches which deal with the construction of counterfactuals internally (i.e. cases without comparison groups). Advanced modelling approaches can contribute to net-impact assessment through an improved consideration of the complexity of public goods and environmental assessments, and explicit consideration of other intervening factors and theoretically-sound counterfactual assessment in situations without available comparison groups (non-participants).

In this section we focus on a few examples of environmental methods tested in climate stability and soil quality case studies which, based on our reviews, have not been used in previous RDP evaluations. These case studies provide examples for advanced environmental methods dealing with the complexity of public goods and environmental assessments and the explicit consideration of other intervening factors. In addition, we identify examples of economic-based models which were tested in climate stability and water quality case studies for their suitability in dealing with situations without comparison groups (e.g. in situations of area-wide uptakes of measures).

The Carbon Footprint (CF) method, tested in the Climate Stability case study in Italy, allows for a robust estimation of the emission based on a well consolidated procedure now also available under ISO rules. CF includes greenhouse gas (GHG) absorption and emission during the life cycle of a product or service, from the extraction of raw materials to its final use. In this way, CF can be considered as a sub-set of data derived from Life Cycle Assessment (LCA). CF can be applied at process level and at farm level with no particular difficulties to estimate the emissions of RDP participants and of the control groups. With sufficiently representative data of the process/farm

samples, micro-level results can be aggregated to provide a robust estimation of macro-level effects. The existence of a well-established farm sample, such as FADN, is a good starting point for the creation of a database for participants and non-participants. But a satisfactory estimation of carbon emissions and sequestrations requires the collection of additional data on farming practices, generally not available in the existing databases, and a significant amount of time for calculating the final carbon footprint. Besides, the application of the method for elaborate statistics-based evaluations of the comparison groups relies on a sufficient number of observations which may increase the whole monitoring and evaluation costs.

The soil quality case study in Hungary tested the application of the Universal Soil Loss Equation (USLE) for modelling soil erosion in combination with the CLUE model (Conversion of Land Use and its Effects) (Verburg et al., 2002). The modelling approach enabled the explicit consideration of other intervening factors influencing soil erosion (sample selection issues) such as rainfall intensity, slope length, slope steepness and land use, which informed the comparison of areas with and without the policy measures. The CLUE model simulates land-use transitions over time and can thus provide a solution for the creation of ‘before and after’ data in the absence of monitoring data. The method requires substantial modelling effort, which might not be feasible for short-term evaluation contracts, in particular as indicator values for different years need to be modelled and analysed separately.

The Scottish soil case study applied the Integrated Valuation of Ecosystem Services and Trade-offs (InVEST) model for the modelling of change in ecosystem services, which is more commonly used for ex-ante assessment, but has proven useful for an ex-post evaluation in data-poor conditions. The USLE equation used for the soil erosion approach is well established as the most effective way to assess rates of soil erosion. It takes account of the importance of the spatial distribution of RDP measures with respect to their impacts on soil erosion, and of the extent of retention of the soil eroded within water sub-catchments. Particular strengths of the modelling approach are the consideration of local environmental characteristics and the establishment of theoretically robust causal relationships. However, the accuracy of the results from the model is dependent upon the level of spatial detail of the input data for the model. Dedicated processes for monitoring soils in relation to different RDP measures would further improve the capability of the modelling approach to contribute to the assessment of net impacts.

*Contributions of tested advanced modelling approaches at micro and macro level: Lack of comparison groups (non-participants)*

The DREMFIA model is an economic modelling approach which was tested in the Finnish climate stability case study for its suitability in dealing with situations without comparison groups, and for its capabilities in taking into account indirect effects such as displacement effects at a macro level. Temporal dimensions of environmental impacts are directly incorporated in the dynamic modelling framework and policy impacts are quantified based on before and after simulations. The modelling framework provides the flexibility to simulate different counterfactual scenarios and the regional

differentiation enables the interpretation of indirect effects at a macro level, such as displacement effects. Care must be taken with respect to the assumptions applied to implementation of the policy measures in the modelling framework to ensure that the causal relationships of the policy measures and related land-management changes are theoretically sound. The complexity of the modelling framework limits its suitability for RDP evaluations to long-term evaluation contracts or the use of already existing models, and requires particular modelling expertise. The application of such a modelling approach for other public good impacts, for example biodiversity wildlife, would rely on the use of proxy indicators directly linked to agricultural land management.

The Finnish water quality case study addressed the control group formation through structural economic modelling. The structural model is used as the counterfactual of non-participation to the agri-environmental programme which is not possible to construct due to the lack of a non-participant control group (90% of farmers participate, covering approximately 95% of Finnish UAA). A biophysical model is used to convert simple pressure indicators (fertiliser use) into more advanced figures of pressure (run-off) using transfer functions from run-off to environmental damage (also in monetary terms). The results are based on a theoretically sound economic model of a representative farm which is calibrated with real-world data. Furthermore, the approach using an environmental impact transfer function provides a robust assessment of environmental impacts. The structural model approach enables counterfactual analysis with missing comparison groups, and is theoretically sound and more robust in comparison to other methods that would rely on naïve approaches to baseline farmer behaviour. However, the results of the case study highlight a few limitations and potential problems in the application. Animal farms are not included in the tested model, as it is under development and a single econometric model may not be able to capture the differences between crop and animal farms. Despite the intuitively clear adaptability of the structural models to impact analysis, problems related to acquiring new FADN data and recalibrating the model to pass consistency checks in the analysis proved surprisingly difficult.

#### *Contributions to the integration of counterfactuals and sample selection issues*

The results of the case studies clearly show that, even in situations with data gaps, at least some sample selection issues can be considered through an ad-hoc approach, e.g. selecting participants and non-participants in close proximity. However, in cases where non-participants do not exist due to the area-wide implementation of measures, or in cases of aggregated macro-level evaluations of programme effects, advanced modelling approaches such as dynamic partial and general equilibrium models provide a theoretically sound alternative for robust before-and-after counterfactual assessments for climate and water quality impacts of RDPs.

The results of the case study highlight possible solutions for the application of elaborate counterfactual evaluation in situations with limited availability of, and access to, data. Applications of advanced statistics-based approaches, such as propensity score matching, with smaller samples and data gaps can still improve the robustness of the results compared to using ad-hoc approaches to deal with sample selection issues. But additional and / or specifically targeted environmental monitoring programmes are needed to fully utilise the potential advanced statistics-based approaches.

<b>Counterfactual integration</b>	
Statistics-based options to deal with sample selection issues	
+ Application with smaller samples and data gaps can still improve the robustness of results compared to using ad-hoc approaches to deal with sample selection issues.	
- Additional and / or specifically targeted environmental monitoring programmes are needed to fully utilise the potential advanced statistics-based approaches.	
Evaluation options without comparison groups	
+ Theoretically sound and more robust approach to deal with area-wide implemented policy measures or counterfactuals at macro level.	
- Time and resource constraints and required modelling expertise might limit the practical application.	
Naïve counterfactual assessments – ad-hoc approaches for sample selection issues	
+ Sample selection issues can, and have, to be considered in naïve approaches through ad-hoc consideration in the design of comparison groups.	
- Contribution to the quantification of net-effects is very limited.	
General key issue	
 Choice of indicator, data availability and the possibilities to construct counterfactual - poor indicator with good counterfactual may not be preferable to a good indicator with a naïve counterfactual approach and comparison group design	

Figure 2 Counterfactual integration

The choice of indicator relates to data availability and the possibilities to construct a counterfactual. Essentially this means that the evaluator may need to prioritise the impact indicators available and see the level of counterfactual analysis possible in each case before choosing the method of constructing the counterfactual (unless more than one approach is used). A poor indicator with a good counterfactual may be preferable to a good indicator with more circumstantial evidence on impact.

In addition, the results of the case studies highlight the importance of the availability of, and access to, environmental monitoring data in combination with key secondary databases. The case studies applied practical solutions to existing data gaps such as the application of national and specific regional and local monitoring programmes from different organisations, the application of freely-available spatial data such as Google Earth and remote-sensing data e.g. Copernicus Programme and a combination of different data sources to enable bigger samples. Negotiations to obtain data access

should start as early as possible in the evaluation process to account for time-consuming processes in the context of different data protection laws.

Data gaps constrain the effectiveness of direct environmental indicators and advanced methods. The performance assessment of the evaluation approaches carried out in the case studies highlights data issues as the single most important factor influencing the effectiveness of the evaluation approaches. The results of the case studies indicate that the cost-effectiveness of monitoring programmes and environmental evaluations can be improved through strategic sampling. More targeted environmental monitoring programmes would facilitate a more robust quantification of deadweight effects and causal relationships and other intervening factors. However, the cooperation and good coordination between monitoring organisations, managing authorities and different ministries needs to be further strengthened.

The implications of scenarios of additional efforts to increase the sample size and to improve the spatial coverage of the monitoring programme as well as of strategic sampling design of monitoring programmes for the cost-effectiveness of evaluations were further analysed in a few selected case studies.

#### ***Possible solutions for dealing with data gaps – impact on the cost-effectiveness of evaluation approaches***

The need for improvements of the data environment is fundamental to facilitate the application of advanced methodological approaches. However, the impacts of data gaps on the effectiveness of indicators and methods need to be compared with the additional cost of improved environmental monitoring programmes. This requires the consideration of different scenarios for future environmental monitoring programmes.

Four case studies (water quality- Germany, climate stability - Italy, biodiversity wildlife – Hungary and landscape - Scotland) were selected to develop cost scenarios in order to show ways to optimise the resource use and facilitate the application of the tested evaluation approaches. Those scenarios show possibilities how to improve the cost-effectiveness of the tested evaluation approaches by changes in the data environment or access. The expected impacts on the related costs and on the performance of the evaluation approach were analysed in WP7.

Scenarios in both case studies tested the availability of additional survey or monitoring data and impacts of reviewing or introducing strategic sampling targeted at the needs of impact evaluations of RDPs. The strategic sampling approach improves the coverage of participants and non-participants and reduces the selection bias, which leads to a more robust net-impact assessment. The new CMES requires the evaluation of synergies and conflicts between measures and focus areas, which is important evidence for recommendations on particular territorial priorities in future RDPs. The strategic sampling approach enables integration of different combinations of measures, and analysis of synergies of combined implementation of measures under the same, or between different, focus

areas. Moreover, a strategic sampling approach and an increased sample size improve the representativeness of the data and the compatibility with local environmental and farm structural data which facilitates the upscaling of the results to the whole programme area.

In most cases the adaptation of a strategic sampling approach for environmental monitoring data for RDP evaluation purposes leads to increased sample sizes compared to the status quo. However, a review of the strategic sampling approach can also lead to a reduction in sampling sizes of existing monitoring programmes, and thus to cost reductions, for example in cases where particular sub-sets of the sample can be reduced without constraining the impact evaluation. Through the integration of multiple time periods, panel data can be created and elaborate-statistics evaluation methods applied, e.g. Propensity Score Matching combined with a Difference in Difference approach. The clearer attribution of environmental changes to the implemented measures and programmes enables more robust recommendations to improve the effectiveness of RDPs.

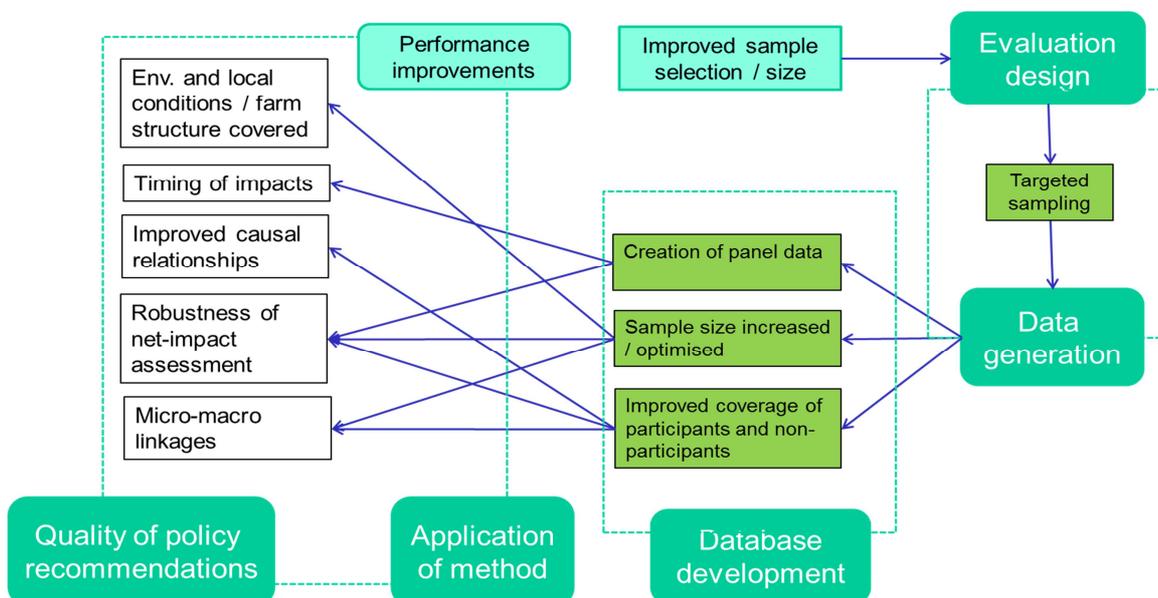


Figure 3 Impact of improved environmental monitoring programmes on the performance of evaluations

These improvements in the effectiveness of environmental impact evaluations cost surprisingly little, at least if one puts the additional cost into the context of the overall RDP budget. The tested examples show that in some cases those improvements can be achieved with a small increase in cost. For example, the revisions to the strategic sampling applied to existing water quality data in Lower Saxony in Germany to increase the effectiveness of RDP evaluations of water quality impacts only resulted in an increase of 2 percent in monitoring costs. Also, small efforts such as the integration of alternative existing data sets or a more detailed analysis and processing of available data can already improve the effectiveness of evaluations. Further cost savings can be achieved by embedding additional data collection, or more generally, environmental monitoring for the evaluations of RDPs into a multi-purpose monitoring system.

Whether the developed scenarios and their results are transferable to other cases requires further validation. The transferability for indicators that are applied across member states (e.g. the farmland bird index) is probably higher than for country-specific situations. However, the improvements of the different scenarios show ways of enhancing data quality and/or quality which are expected to be useful for monitoring data for varying indicators or methods. A number of lessons can be derived for future environmental monitoring programmes:

- Setting data prerequisites at the beginning of each programming period leads to sound statistical analyses of environmental impacts and robust recommendations
- Planning of impact evaluations at the stage of scheme design ensures necessary data availability for consistent evaluation
- Adjustments to sampling and monitoring methods targeted at RDP evaluation improves cost-effectiveness of the evaluation process
- Embedding additional data collections for improving RDP evaluations into a multi-purpose monitoring system eventually leads to public resource savings and more comprehensive data sets.

### *Developing the methodological framework*

The experiences from the case study testing and cost-effectiveness assessment informed the development of the logic model based methodological framework for environmental evaluations of RDPs. The framework builds on step-by-step guidance in designing cost-effective evaluation approaches provided by the logic models.

The evaluation of RDP impact on the environment consists of three main components: a sound counterfactual design, and assessments at micro and macro levels. These three elements of the framework (shown in simplified form in Figure 4) are linked and, following consistency checks between micro and macro levels, they collectively inform the net impacts of RDP. The framework provides a context and transparency that assists in structuring the assessment by defining sound comparison groups, and includes a check that the results from both micro and macro level are consistent, which can improve the quality of the evaluation.

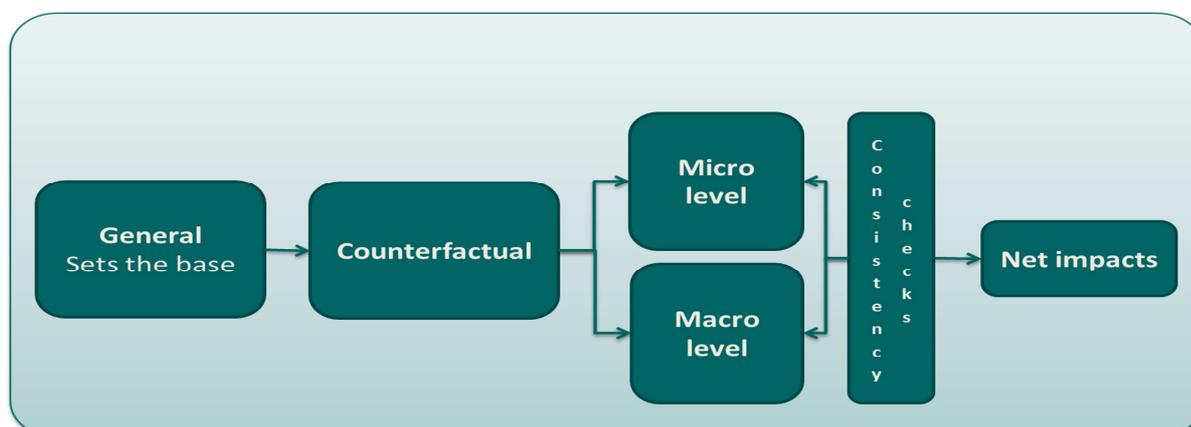


Figure 4 Simplified Evaluation Framework showing the elements of counterfactual, micro and macro levels

The first general layer of the methodological framework (logic models) provides an overview of the overall intervention logic and structure suggested to pursue in the evaluations of environmental RDP impacts (from the decision to evaluate specific measures or the whole programme, the selection of specific indicators to the integration or dis- and aggregation of micro and macro level results to a consistent net impact assessment). The general layer includes the following steps:

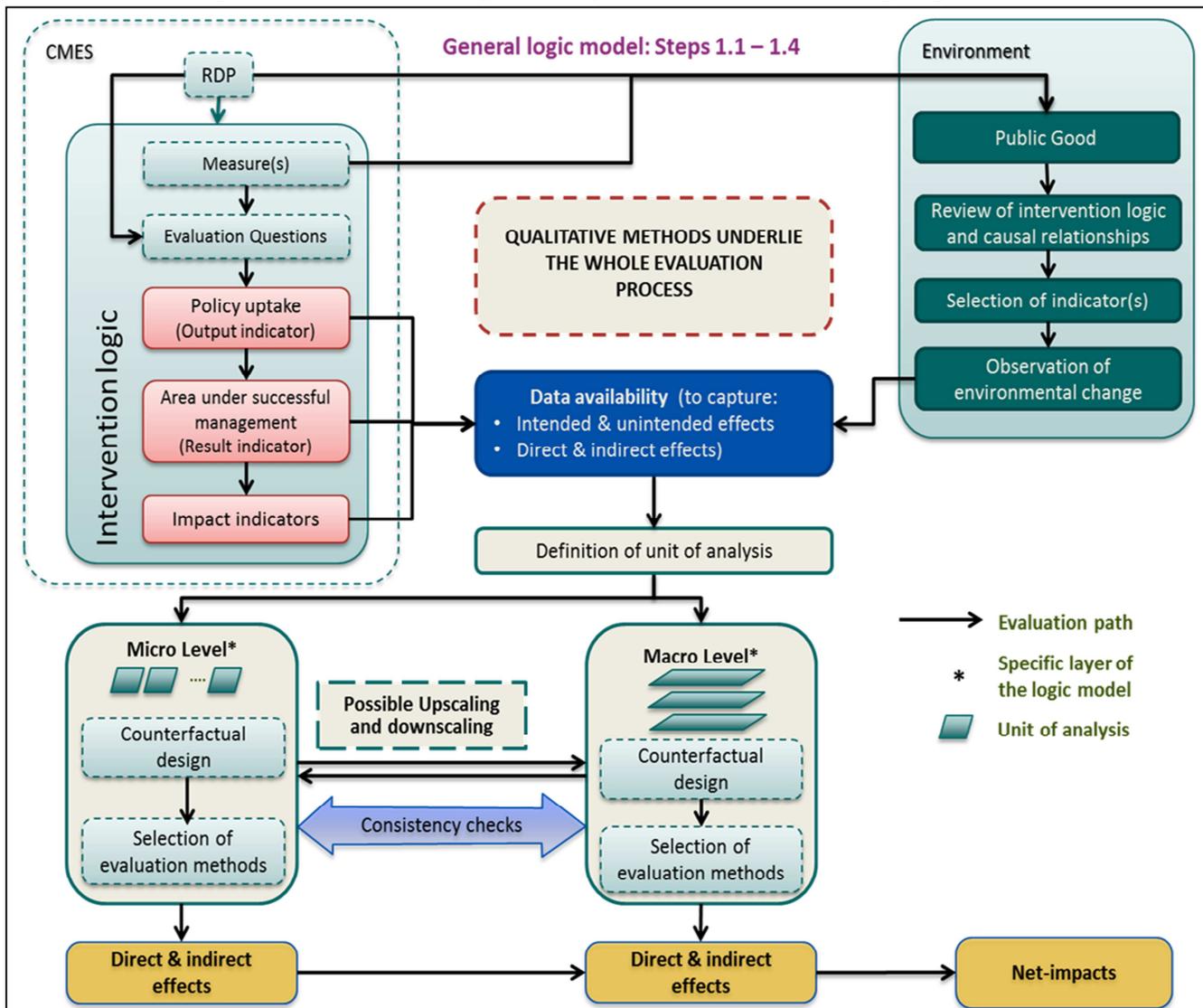


Figure 5 Overview of the general layer of the methodological framework for environmental evaluations of RDPs

Step 1.1: Applying the CMES intervention logic

Step 1.2: Selecting indicators for public goods

Step 1.3: Definition of unit of analysis

Step 1.4: Counterfactual design of micro and / or macro level evaluations

A critical part of developing cost-effective evaluation approaches is the development of the counterfactual design. The following steps describe the work flow of the logic model for designing one or more counterfactuals. It is applicable for both micro and macro level evaluations. The logic

model highlights the importance of defining and identifying comparison groups from available data. The formation of comparison groups is particularly important when self-selection to of programme participation is likely. When farmers are not randomly assigned as participants to the evaluated programme, a simple comparison of programme participants and non-participants may lead to biased impact estimation of an unknown magnitude and direction. The logic model considers the identification of comparison groups predominantly from a data perspective. An explicit process categorising the possible methods to design a counterfactual with available data is important even if data are lacking. The logic model can guide the evaluator towards new approaches, better planning of future data gathering, and also serve as an initial thought process for methods that are less reliant on data availability. The counterfactual layer includes the following steps leading to three different counterfactual based evaluation options:

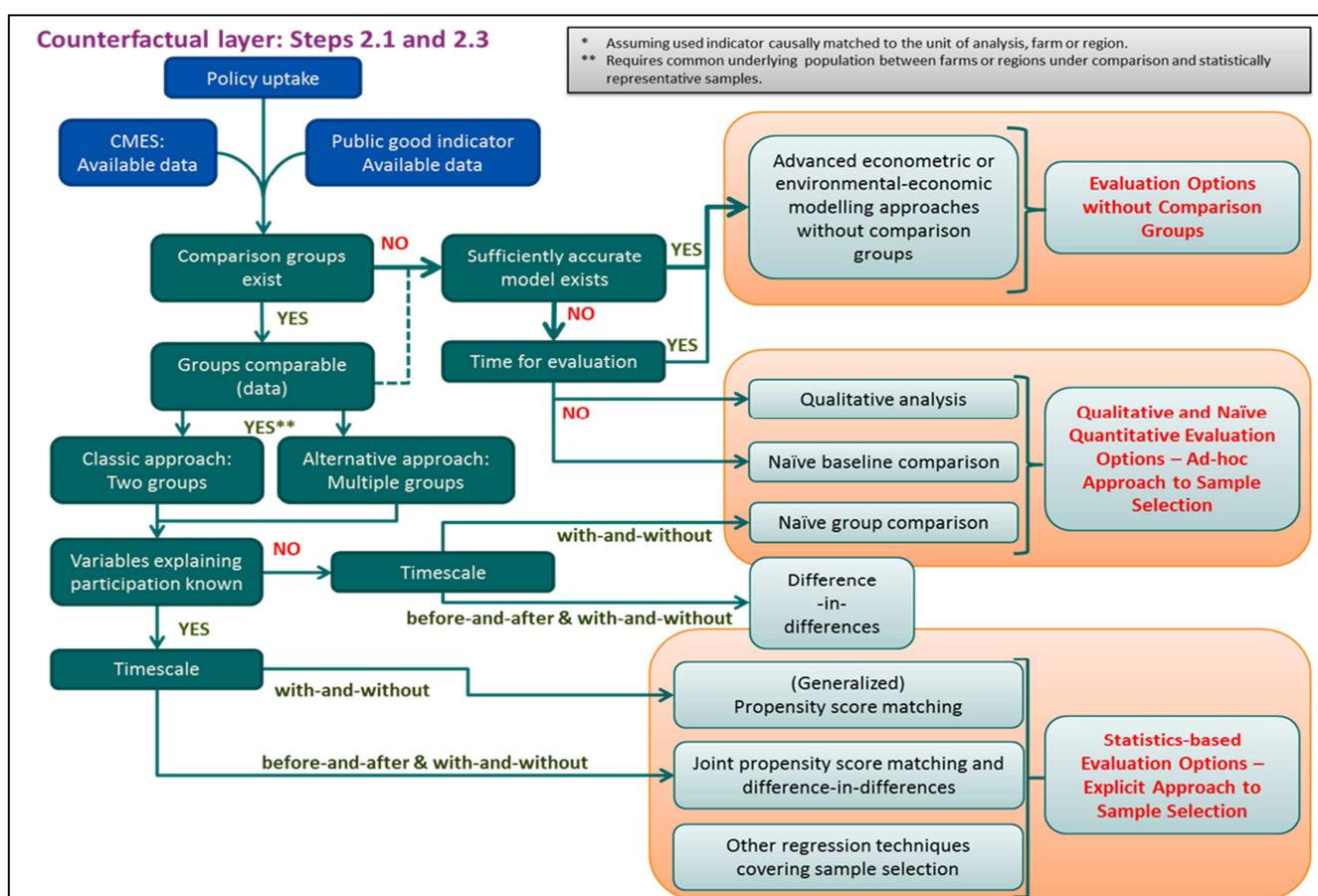


Figure 6 Steps in counterfactual design

Step 2.1: Inputs to designing a Counterfactual

Step 2.2: Defining comparison groups

Step 2.3: Choice of evaluation options

- a. Evaluation options without comparison groups
- b. Qualitative and naïve quantitative evaluation options
- c. Statistics-based evaluation options

The workflow for the micro-level logic models leads the reader to methods which contribute to a consistent assessment of net impacts at micro and macro levels. For each of the three possible counterfactual designs, an individual micro-level logic model has been created. The first two steps of these logic models are the same. The third step (Step 3.3) reflects the three counterfactual-based evaluation options and leads to different micro-level methods:

Step 3.1: Definition of the unit of analysis and consistency of selected indicators

Step 3.2: Assessment of data quality

Step 3.3: Selection of counterfactual approach

Step 3.3a: Evaluation options without comparison groups

- Example methods include: Structural models, integrated models and agent-based modelling

Step 3.3b: Qualitative and Naïve Quantitative Evaluation Options – ad hoc approach to sample selection

- Example methods include: Sustainability indicators, Ecological footprint and composite sustainability indicators, integrated models and agent-based modelling.

Step 3.3c: Statistics-based evaluation options

- Example methods include: Matching methods such as propensity score matching (PSM), Difference in Differences and regression discontinuity design.

Step 3.4: Micro-Macro consistency and validation

The methodological framework for macro-level evaluation follows the same principal structure and has three different macro-level logic models, one for each of the three types of counterfactual designs. The steps within the three macro-level logic models are the same:

Step 4.1: Definition of the unit of analysis and consistency of selected indicators

Step 4.2: Creation of consistent spatial data

Step 4.3: Selection of counterfactual approach

Step 4.3a: Evaluation options without comparison groups

- Example methods include: CGE and PE modelling frameworks, spatial econometrics, landscape metrics

Step 4.3b: Qualitative and Naïve Quantitative Evaluation Options – ad hoc approach to sample selection

- Example methods include: Ecological footprint, multi-criteria assessments and multifunctional zoning.

Step 4.3c: Statistics-based evaluation options

- Example methods include: Spatial econometrics and landscape metrics.

Step 4.4: Net Impact Assessment

Step 4.5: Micro-macro consistency

The triangulation of different methods and approaches used to assess impacts at micro and macro levels can be used to validate the consistency of results at these levels. The results of macro-level impacts, based on the upscaling of micro-level results, can be compared with macro-level impacts based on the application of a specific macro-level method or approach (e.g. a macro-level modelling approach or a specific calculation of indicators at macro level). For example, the up-scaling of assessments of the impacts of rural development measures on water quality using the indicator GNB at a farm level (i.e. micro level) can be compared to the results of an assessment of GNB at a catchment level. However, the combination of a bottom-up approach based on evaluations of the RD measures at micro level (followed by an upscaling of results) and a top-down approach using a specific macro-level method to assess RDP impacts is only recommended with an appropriate level of resourcing and longer-term evaluation contracts. More details of the methodological framework is available on the project website ([www.envieval.eu](http://www.envieval.eu)) and in the handbook for environmental evaluations (Deliverable D9.5)

***Integration of cost-effectiveness aspects in the methodological framework (logic models)***

Each step of the logic model (methodological framework) consists of several activities where decisions on the development and implementation of the evaluation exercise have to be taken. The integration of cost-effectiveness aspects into the developed methodological framework was done in WP7. The following figure shows the five steps of the evaluation cycle and the decisions that influence the evaluation design and thus the cost and effectiveness of the approach.

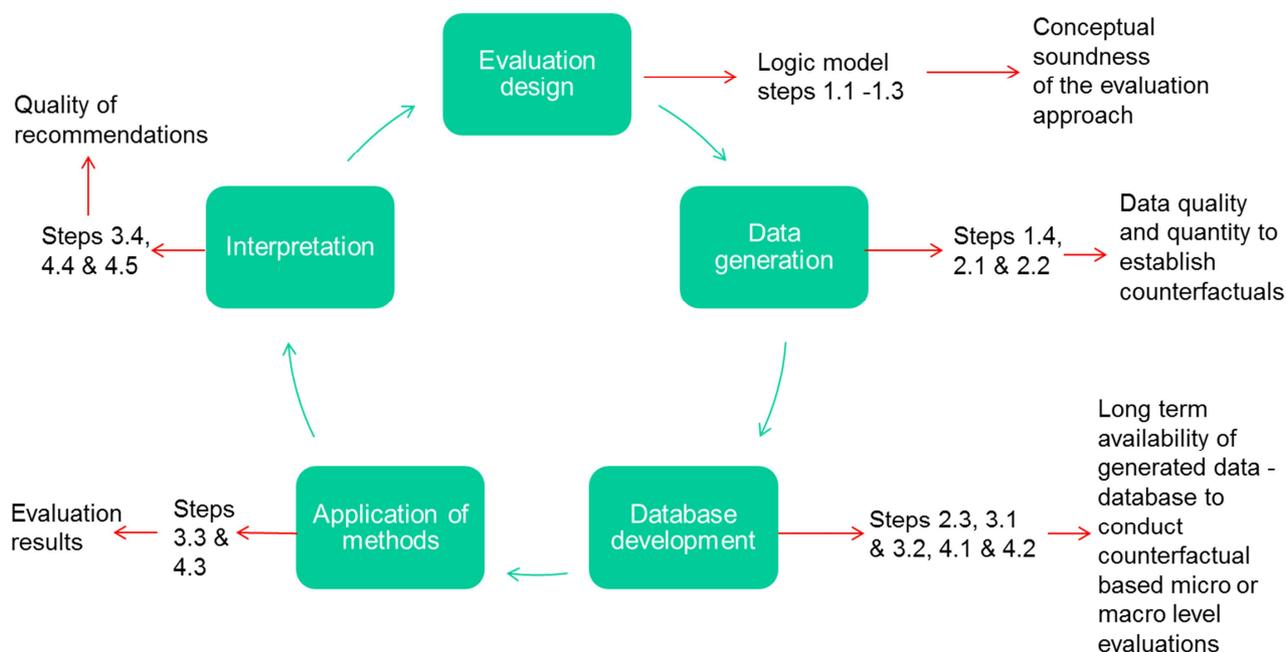


Figure 7 Evaluation cycle with the impacts on effectiveness

In the first step, the evaluation design, several decisions according to the conceptual design of the evaluation approach have to be taken. The application of the CMES logic framework has to be conducted at the beginning of the evaluation design (Step 1.1). Additional indicators have to be

selected if necessary (Step 1.2). This can be associated with high cost as it requires additional work to the application of the CMES. However, the selection of suitable additional indicators could increase the effectiveness of the evaluation exercise and might be beneficial. Further, data requirements and available data need to be reviewed. These activities are crucial for the successful application of the statistical analysis and have a strong impact on the effectiveness. High cost savings are possible if existing data sources could be used, as data collection is usually more expensive. Conceptual decisions have further to be drawn on the selection of functional units (Step 1.3).

It can be concluded that decisions at this evaluation stage are mainly associated with increased labour cost as more time is spent on the development of the evaluation approach. Decisions at that stage then influence the effectiveness of the evaluation approach at all other stages of the evaluation. The right decisions in the beginning of the evaluation process are essential for the successful application of the evaluation method and thus are usually worth the higher cost.

The second evaluation phase is associated with data generation activities and includes tasks related to the use of existing data sources and the collection of additional primary data, if necessary. Data is assessed to enable statistical analysis with counterfactual design at a micro and macro level to enable net-impact assessments. Data availability for counterfactuals needs to be checked (Step 2.1) as well as the possibilities to construct robust counterfactuals with or without comparison groups with the existing data (Step 2.2). If additional primary data collection is conducted, this evaluation step can be at a high cost. The mode of data collection and the sampling strategy have a high impact on the effectiveness of the evaluation as this provides the basis for a sound statistical analysis. The use of existing data sources is usually associated with lower cost as most evaluators have access to a variety of free data sources. Monitoring data are often not directly targeted for use for evaluation purposes and often does not meet the needs of evaluation. This has a strong negative impact on the effectiveness of the evaluation as the results are not robust or the statistical analysis does not cover all aspects of rural development impacts. Thus, increased efforts in planning and design of data collection are worth the improved sampling or coverage of rural development impacts despite the higher labour costs.

In the third phase of the database development and maintenance, important decisions have to be taken which influence the cost-effectiveness of the evaluation approach. The evaluation option for counterfactual based analysis (Evaluation Options without Comparison Groups, Qualitative and Naïve Quantitative Evaluation Options or Statistics-based Evaluation Options) depending on the existing data availability is selected (Step 2.3). Decisions are related to development of the database to conduct counterfactual based micro (Step 3.1 and 3.2) and macro (Step 4.1 and 4.2) level evaluations. Activities include the set-up of data infrastructure for counterfactuals and development of procedures and protocols. Decisions relating to these activities have a strong impact on the effectiveness of the available data sources. Further, the maintenance of the database is important for ensuring the long-term availability of data generated. Decisions in this evaluation phase are mainly

related to increased work load or the kind of equipment (e.g. software) that is used in the analysis. The investment in development of a robust database and its maintenance could increase the effectiveness of the evaluation method and enable the use of the data base for future evaluations.

The application of the method (fourth phase of the evaluation cycle) uses the database developed to implement counterfactual based micro and macro-level analysis (Step 3.3 and 4.3). Analysis is based on the previous assessment. The suitability of the selected indicators based on the data availability is tested and adaptations are implemented (if required). Decisions are required about the mode of analysis and variations of the testing which directly influence the quality of the evaluation results. Usually, this decision is related to an increased work load for the evaluator. The accuracy and quality of the analysis is directly influenced by the decisions in this evaluation step.

The final phase of the evaluation cycle refers to the Interpretation of results and conducting consistency checks (Step 3.4 and 4.4). The results of the analysis need to be communicated to the target group. Depending on the complexity of the analysis greater efforts could be required to 'translate' scientific results into understandable and unambiguous policy recommendations. Decisions are required regarding time spent for the evaluation, usually with associated investment in personnel and equipment, but innovation may offset those costs, such as in relation to communicating results.

Conducting consistency checks (Step 3.4 and 4.4) is essential to validate the results of the analysis and increase its robustness. Decisions have to be made on the mode of analysis for consistency checks. Costs arise due to increased staff time on consistency checking. Further, additional costs for equipment might be necessary, e.g. when the use of further statistical software is required. The quality of the results increases when sufficient time is spent on the communication and development of policy recommendations as well as the validation of the results through consistency checks. Thus, decisions in this evaluation step have a strong impact on the effectiveness of the evaluation approach.

In conclusion, in all evaluation phases decisions are required which will influence the cost-effectiveness of the evaluation approaches. This is particularly true of decisions at the outset of the evaluation cycle, thus in the first steps in the application of the logic model, which have impacts on the overall effectiveness of the evaluation as they influence data generation, database development and applications of the evaluation method. However, good decisions at the outset (e.g. with respect to selection of indicators in Steps 1.1 and 1.2) cannot support good quality evaluation results if subsequent decisions in the evaluation process (e.g. with respect to the selection of counterfactual options in Step 2.3) inhibit the analysis. Thus, an appropriate level of resources can be expected to facilitate a successful evaluation.

### ***Fact sheets for the handbook***

The fact sheets developed in WP8 are a final outcome presenting a short summary of the main characteristics of the indicators and methods tested in ENVIEVAL. They provide information on

why and for which policy aspects the indicators or methods can be used, and where the required data can be sourced and obtained. The fact sheets summarise the strengths and weaknesses of the indicators and methods, and highlight their contribution to addressing the main challenges. An adjusted ‘SWOT’ framework is used to synthesise the key advantages, disadvantages and contributions of the indicator / method.

The general structure of the indicator and method fact sheets is as follows:

Indicator fact sheets:

1. Definition / description of the indicator, including environmental public good, type of indicator, reflected RDP priority and focus area, unit of measurement, type of data required and scale and level of application
2. Existing data sources including EU, member states and regional databases
3. Context of the case study testing, including case study area, policy context, used data and evaluation approach tested
4. Strengths and weaknesses of the indicator
5. Recommended application

Method fact sheets:

1. Definition / description of the method, including the environmental public good, type of method, micro or macro level application
2. General requirements including data requirements and skill requirements
3. Consideration of counterfactuals
4. Context of the case study testing, including case study area, policy context, used data and evaluation approach tested
5. Strengths and weaknesses of the method
6. Recommended application

The indicator fact sheets focus on additional non-CMES indicator tested in the ENVIEVAL project for their contributions to address indicator gaps in environmental evaluations of RDPs. The method fact sheets focus on advanced modelling approaches tested at micro and macro level for dealing with the complexity of public goods, considering other intervening factors and providing solutions for situations without (or very limited) non-participants. The fact sheets were reviewed by the stakeholder reference group and their comments and feedback integrated in the final version of the fact sheets.

#### *d. Dissemination and Impact*

##### *Dissemination*

Practical application of the results and outputs of ENVIEVAL such as the methodological framework and handbook for environmental evaluations of RDPs depends highly on its acceptance by concerned evaluators, managing authorities and EU level stakeholders, its relevance in addressing national and EU policy interests and data availability and organisation of public administration structures. Stakeholder engagement was implemented as a two-way process of communication between the project team and the stakeholder groups, with stakeholder involvement throughout the research cycle. As these aspects closely deal with the “human factor”, stakeholder engagement was understood as a very important component of the project to maximise acceptance of the planned results. A full work package was devoted to stakeholder involvement and dissemination (WP9).

The majority of project stakeholders have already been enlisted into the project application. However, at further stakeholders at national levels joint the project at the beginning taking into consideration national circumstances and institutional structures. A simplified structure (used also in the Water Framework Directive) of stakeholder engagement was applied:

- Information. Stakeholders are provided with or have access to information.
- Consultation. The views of stakeholders are sought, real interactions and feedback takes place.
- Active involvement. Stakeholders share decision-taking powers.

Table 4 List of stakeholders structured according to the intensity of involvement

<b>Involvement level</b>	<b>Stakeholder groups</b>	<b>Involved tools/methods</b>
Information	European and national farmers’ associations; international and national environmental NGOs; national RDP developers and decision-makers; business associations related to rural development.	Project website; Fact sheets; Policy briefs; Handbook; Final Conference
Consultation	EU and national evaluation experts; Managing Authorities of RDPs; environmental monitoring experts; European Evaluation Helpdesk; representatives of DG Agri, DG ENV.	Stakeholder Reference Group; International stakeholder workshops; National meetings and workshops.
Active involvement	Project desk officer in EC, science community related to RDP evaluation, indicator definition, public goods assessment, public administration; agricultural policy experts.	Project Advisory Group; Direct personal approach; Project reporting and partners meetings.

The main consultations with the different types of stakeholders are summarised in Table 5.

Table 5 Overview of national and international stakeholder consultations

Year	Level	Objectives	Participants	Mechanisms
2013	National (12 countries)	<ul style="list-style-type: none"> <li>• identify key gaps &amp; problems in environmental evaluations from the stakeholders point of view</li> <li>• collate information on why certain indicators, data bases and methods have been used in past evaluations</li> <li>• assess the expectations and requirements for improved and alternative indicators and methods</li> </ul>	Evaluators, managing authorities, monitoring organisations, European Evaluation Helpdesk	Interviews with guideline-based questionnaire
2013	International	<ul style="list-style-type: none"> <li>• review strengths and weaknesses of current indicators and evaluation methods</li> <li>• reflect on stakeholder expectations and requirements for improved and alternative indicators and methods</li> <li>• identify stakeholder priorities for the case study testing</li> <li>• identify key stakeholder requirements for future evaluation methods.</li> </ul>	Evaluators, managing authorities, monitoring organisations, European Evaluation Helpdesk	Workshop
2014	National (7 countries)	<ul style="list-style-type: none"> <li>• validate' the policy and stakeholder relevance of the selected case study regions with the wider group of expert and officials</li> <li>• fill knowledge gaps and agree on further collaboration in the process</li> </ul>	Evaluators, managing authorities, monitoring organisations	Workshops and interviews
2014	International	<ul style="list-style-type: none"> <li>• review the applicability of draft logic models for the design of evaluation approaches</li> <li>• review the stakeholder relevance of the cost-effectiveness assessment</li> <li>• discuss the practical relevance of the case studies in the context of the forthcoming ex-post evaluations and future evaluation tasks in the period 2014-2020</li> <li>• identify key stakeholder requirements for the forthcoming ex-post evaluation and the potential contributions of the ENVIEVAL project to methodological guidance</li> </ul>	Evaluators, managing authorities, monitoring organisations, European Evaluation Helpdesk, European Commission	Workshop
2015	National	<ul style="list-style-type: none"> <li>• review the applicability of the revised methodological framework for designing evaluation approaches</li> <li>• review progress of case study testing of indicators &amp; methods</li> </ul>	Evaluators, managing authorities, monitoring organisations,	Interviews and workshops
2015	International	<ul style="list-style-type: none"> <li>• validate a comparative assessment of the performance (effectiveness) of the evaluation approaches</li> <li>• discuss the practical relevance of the case studies in the context of forthcoming evaluation tasks in 2014 – 2020</li> <li>• review the structure of the methodological handbook</li> </ul>	Evaluators, managing authorities, monitoring organisations, European Evaluation Helpdesk	Workshop
2015	National	<ul style="list-style-type: none"> <li>• validate stakeholder relevance of the fact sheets of the tested indicators and methods</li> <li>• validate the applicability of the cost-effectiveness assessment of the scenarios for future env. monitoring programmes</li> </ul>	Evaluators, managing authorities, monitoring organisations,	Interviews and workshops
2015	International	<ul style="list-style-type: none"> <li>• discuss indicators and methodological solutions to main challenges in environmental evaluations</li> <li>• discuss data gaps and the cost-effectiveness of improved environmental monitoring programmes for RDP evaluations</li> <li>• discuss key elements of the methodological handbook for environmental evaluations of RDPs</li> </ul>	Evaluators, managing authorities, monitoring organisations, European Evaluation Helpdesk, European Commission, Scientists	Conference

In addition, to the main consultations ongoing bilateral discussions between project team members and members of the project advisory group and stakeholder reference group took place on a regular basis. Particular attention was paid to dissemination activities with key stakeholders at EU level such as DG Agri and the European Evaluation Helpdesk.

**Key EU level dissemination activities:**

Introduction of project to different units of DG Agri, Brussels, 19<sup>th</sup> February 2013:

- 15 participants from different units of DG Agri and OECD

Good Practice Workshop with European Evaluation Helpdesk on ‘Assessing Environmental Effects of Rural Development Programmes: Practical solutions for the ex post evaluation 2007-2013’, Vilnius, 27<sup>th</sup> and 28<sup>th</sup> of October 2015:

- 40 participants of evaluators, managing authorities and scientist from across the EU

Presentation of ENVIEVAL results at the 8th Meeting of the Expert Group on Monitoring and Evaluating the CAP, Brussels, 12th November 2015:

- 60 participants of DG Agri, evaluators and managing authorities from across the EU

Lunchtime Seminar at DG Agri on ‘Evaluation of environmental effects of RDPs - key results and recommendations from the ENVIEVAL project’, Brussels, 20<sup>th</sup> of November

- 25 participants from different units of DG Agri and DG ENV

Utilisation of multiplier effects through dissemination of ENVIEVAL results through the European Evaluation Helpdesk:

- Dissemination of presentation on the website of the Helpdesk (e.g. <https://enrd.ec.europa.eu/en/evaluation/european-evaluation-helpdesk-rural-development/good-practice-workshops>)
- Short articles on the project in the newsletters of the Helpdesk (e.g. Rural Evaluation News No. 2 [http://enrd.ec.europa.eu/sites/enrd/files/newsletter\\_2-en\\_2502\\_0.pdf](http://enrd.ec.europa.eu/sites/enrd/files/newsletter_2-en_2502_0.pdf))

Further emphasis of the dissemination activities was on scientific and combined scientific and stakeholder events. Examples include presentations of the ENVIEVAL project at the EAAE 2014 Congress ‘Agri-Food and Rural Innovations for Healthier Societies’ in Ljubljana, Slovenia, in August 2014 and SRUC-SEPA conference on ‘Delivering Multiple Benefits from our land: Sustainable Development in Practice’ in Edinburgh in April 2014. One of the main scientific dissemination events was the organisation of two sessions at the Annual Meeting of the American Geographers in Chicago in April 2015 to expand the outreach of the project beyond the EU boundaries. The organised sessions combined presentations from the ENVIEVAL project with

presentations from other relevant EU projects such as the SPARD project, as well as presentations from other international projects such as the CADWAGO project.

The exchange and cooperation with other EU projects will be continued in the future, with joint sessions organised with the PROVIDE and PEGASUS projects at the conference of the Italian Association of Agricultural and Applied Economics (AIEAA) in June 2016. The topic of the session is 'How to evaluate the environmental impact of Rural Development Programme. Methodological challenges in multiscale and multilevel contexts'. In addition further future dissemination activities are organised at the SRUC / SEPA conference on 'What Future for our Farming Systems? Environmental Challenges and Integrated Solutions' in March 2016 and Annual Meeting of the American Geographers in San Francisco in April 2016.

Final dissemination of project results and outcomes is through policy briefs and the dissemination of the methodological handbook for environmental evaluations. The dissemination is facilitated through collaboration with the European Evaluation Helpdesk and further updating and maintenance of the ENVIEVAL website.

The ENVIEVAL Handbook aims to be a practical guide to help with developing an approach to the evaluation of the environmental impacts of rural development measures and programmes. It presents a logical approach to the design of an evaluation, identifying appropriate methods based on consideration of the requirements, data availability, quality and type.

The Handbook structure:

- (a) the contemporary policy context for an evaluation
- (b) an introduction to the conceptual framework (logic model) and process for designing an evaluation, with a flow chart for each step in the process
- (c) selection of counterfactual for a rural development measure or programme, with identification of methods for a counterfactual analysis best suited to the circumstances
- (d) selection of a suitable method for conducting the evaluation
- (e) identification of appropriate moments to test for consistency between evaluations at micro- and macro-levels
- (f) identification of data related limitations and issues
- (g) working through a logic model
- (h) examples of the application of the framework
- (i) factsheets of the tested indicators and methods

In addition to a policy brief highlighting the general benefits of the logic model based methodological framework, five policy briefs have been produced highlighting a range of key messages and lesson learnt from the ENVIEVAL project.

Table 6 Policy briefs and lessons learnt

Policy brief title	Key messages and lessons learnt
No. 2: Constructing counterfactuals in RDP environmental impact evaluation	<ul style="list-style-type: none"> <li>• Robust counterfactuals are critical for clear attribution of observed changes to policy measures</li> <li>• Counterfactual analysis ensures transparency in impact assessment</li> <li>• Monitoring and farm-specific data need to be collected from non-participants to enable robust counterfactuals</li> <li>• Robust methods are available to assess cases where non-participant data is lacking</li> <li>• Programme level evaluation requires separate top down approach or a qualitative assessment.</li> </ul>
No. 3: Implementation of methodologies for linking farming systems and environmental public goods	<ul style="list-style-type: none"> <li>• Statistics-based approach to counterfactuals at micro level needs well-defined samples with a sufficient number of observations to perform regression models and spatial analysis.</li> <li>• Lack of temporally-differentiated participation data and land cover data are among the main limitations in terms of data requirements</li> <li>• Limited availability of data on farm structure and practice management restricts application of advanced matching techniques.</li> <li>• Proxy indicators are an cost-effective option.</li> </ul>
No. 4: Towards micro-macro level consistency in environmental evaluations of RDPs	<ul style="list-style-type: none"> <li>• The macro-level framework pays particular attention to addressing improved consistency of micro-macro linkages.</li> <li>• The establishment of robust causal relationships and the consistency between micro and macro-level benefits from an integration of RDP monitoring with environmental monitoring programmes.</li> <li>• Combinations of a ‘top-down’ macro-level approach evaluating programme impacts with a ‘bottom-up’ micro-level approach assessing net-effects of different measures or measure combinations increases the robustness and consistency of net-impact assessments.</li> <li>• Plausibility and consistency checks are needed in relation to data used as inputs and the comparison of the results at different levels.</li> <li>• Long-term evaluation contracts and close institutional collaboration on setting a synchronised data-gathering strategy become crucial.</li> </ul>
No. 5: Data issues in environmental impact evaluation of RDPs	<ul style="list-style-type: none"> <li>• The combination and integration of different data sources creates a sufficient number of samples for sound statistical analysis of comparison groups. But particular care must be taken in ensuring the comparability and reliability of different data sets.</li> <li>• Access to aggregated IACS is not sufficient to apply elaborate statistics-based methods. to quantify net effects of RD measures and RDPs. Negotiations to access more detailed IACS data need be started early to account for this time-consuming process.</li> <li>• Early planning of sampling strategy for environmental monitoring considering the specific requirements of RDP evaluations is critical for the evaluation.</li> </ul>
No. 6: Low cost improvements of the environmental monitoring programmes leads to big added value on the effectiveness of RDP evaluations	<ul style="list-style-type: none"> <li>• Setting data pre-requisites at the beginning of each programming period leads to sound statistical analyses of environmental impacts and robust recommendations</li> <li>• Planning of impact evaluations at the stage of scheme design ensures necessary data availability for consistent evaluation</li> <li>• Adjustments to sampling and monitoring methods targeted at RDP evaluation improve cost-effectiveness of the evaluation process.</li> <li>• Embedding additional data collections for improving RDP evaluations into a multi-purpose monitoring system eventually leads to public resource savings and more comprehensive data sets.</li> </ul>

## *Impact*

Agriculture in Europe is important for the provision of a wide range of public goods. Many aspects of the countryside that people value most, and which they expect farming and forestry to provide, are public goods such as biodiversity or landscapes. Agriculture can also help to provide other environmental public goods such as a high quality of air, soil and water, a stable climate and animal welfare. Whilst all types of farming can provide public goods if the land is managed appropriately, there are significant differences in the type and amount of public goods that can be provided by different types of farms and farming systems in Europe. Due to the diversity of agricultural activities and rural environment across the EU, the needs, characteristics and impacts of EU rural development policies vary between different countries and regions. This highlights the necessity of having a flexible methodological framework which combines evaluation tools which can monitor and evaluate the environmental consequences of such policies taking into account specific characteristics and requirements at territorial level. ENVIEVAL has developed a flexible methodological framework for designing evaluation approaches which address the need for rural policy evaluation at disaggregated territorial level to improve the targeting of rural development measures and programmes.

While animal welfare is an objective of the EU rural development policy and a large number of Member States have implemented measures addressing animal welfare (e.g. through animal welfare programmes – measure 215, or farm investment measures – measure 121), there are no real indicators in the CMEF for assessing the efficiency of the measures or the current level of animal welfare. Since animal welfare is of high priority and concern to EU citizens, it is vital to develop approaches to assess changes in animal welfare due to rural development programmes. The incorporation of a suite of outcome-based welfare indicators remains an important target. ENVIEVAL has built on the development of standardised ways to assess animal welfare in recent and current international projects taking into account animal-based measures and resource or management-based characteristics. ENVIEVAL provided recommendations for the integration of outcome-based indicators, taking into account their particular monitoring requirements, into the assessment of the impacts of rural development programmes on animal welfare.

Member States are required to assess the progress, efficiency and effectiveness of rural development programmes (2007 to 2013) in relation to their objectives by means of indicators relating to the baseline situation as well as to the financial execution, outputs, results and impact of the programmes. The Common Monitoring and Evaluation System (CMES), drawn up by the Commission in collaboration with the Member States, provides a single system for the monitoring and evaluation of all rural development measures. A common set of input, output, result, impact and baseline indicators for the rural development programmes is defined.

However, gaps in existing monitoring activities and data availability create difficulties in linking measure-specific evaluation results at a micro level with impact indicators at programme and macro levels. What is lacking is a flexible data and monitoring framework which provides connections

between impacts at different scales linking micro- and macro-level data, including protocols for the aggregation of micro-level data. ENVIEVAL has provided concrete recommendations for dealing with data gaps in environmental evaluations, and has tested the implications of scenarios of improved environmental monitoring programmes for the cost-effectiveness of RDP evaluations at micro- and macro-levels as a basis for the assessment of future performance targets in the new rural development programmes in 2014.

ENVIEVAL has utilised new research opportunities resulting from recent methodological developments to address key evaluation challenges such as counterfactual development, the quantification of net effects and a better understanding of why certain impacts and changes have taken place and how the policies operate. To achieve these improvements, new developments in terms of scenario development, environmental planning tools at farm scale, new spatially explicit bio-physical and land use models, new development in spatial econometrics, farm region methods such as Footprint assessment and Life Cycle Analysis, top down macro-level modelling tools such as CAPRI and RAUMIS, and greater integration of qualitative methods, e.g. through mixed methods case study approaches, have been reviewed and considered. Particular contributions of the different methods tested to netting out deadweight and substitution effects have increased the cost-effectiveness of RDP evaluation in the EU.

The policy relevance of the project outcome has been ensured through close collaboration with the European Evaluation Helpdesk and evaluators and managing authorities by each partner. ENVIEVAL has put an emphasis on stakeholder involvement and user friendliness of the evaluation tools. Several stakeholder consultations (see Table 5) and continuous stakeholder involvement throughout the project has facilitated application of the logic models and methodological handbook by evaluators in the Member States across the EU, and has raised awareness amongst policy makers at national and EU level to further improve causal relationships between environmental characteristics, needs, impacts and expenditures of rural development measures.

The national and regional impact of ENVIEVAL differs depending on specific circumstances in the various countries and regions. In a number of cases, a lack of experience with advanced evaluation methods, development of spatial data infrastructure and the implementation of monitoring frameworks have been reported in the stakeholder consultations. In those, the use of the methodological handbook will be particularly valuable to ensure that consistent and cost-effective evaluation approaches are used in future evaluation tasks in the enhanced Air 2017 and 2019. ENVIEVAL provides methodological guidance on the development and application of indicator and monitoring frameworks, database infrastructure and the design of cost-effective evaluation methods integrating micro and macro-level results and improve the methodological skills of evaluators in the in this programme period.

Discussions within the EC and European Parliament recognise the importance of a ‘strong common policy structured around its two pillars’. One of the main challenges facing EU agriculture post 2020

remains to respond to growing environmental and sustainability requirements; and to contribute to reasonable living standards for primary producers who will have to cope with volatility in markets, including that arising from the effects of climate change. But these challenges arise partly because agriculture continues to undermine its own sustainability by degrading natural capital – pollinators, soil fertility, biodiversity, water and air quality. Making food production more efficient in its use of resources and more viable for the future whilst at the same delivering environmental benefits should be core functions of agriculture beyond 2020. The outputs from ENVIEVAL help to improve the evidence base of the types and net effects of environmental impacts caused by RDPs.

#### **e. References**

European Commission (2005) Agri-environment Measures. Overview on General Principles, Types of Measures and Application. Directorate General for Agriculture and Rural Development. Unit G-4 - Evaluation of Measures applied to Agriculture, Studies. March 2005.

European Commission (2009) Treaty of Lisbon. <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=URISERV%3Aai0033>

European Evaluation Helpdesk (2015) Guidelines - Establishing and implementing the evaluation plan of 2014-2020 RDPs. [https://enrd.ec.europa.eu/sites/enrd/files/uploaded-files/twg-05-ep-june2015\\_0.pdf](https://enrd.ec.europa.eu/sites/enrd/files/uploaded-files/twg-05-ep-june2015_0.pdf)

European Network for Rural Development (ENRD) (2011) Public goods and rural development. <https://enrd.ec.europa.eu/sites/enrd/files/fms/pdf/26065290-F2AE-D77F-BDE0-4DF8472225BE.pdf>

Institute for European Environmental Policy (IEEP) (2010) The provision of public goods through agriculture in the European Union. <http://www.ieep.eu/publications/2010/01/the-provision-of-public-goods-through-agriculture-in-the-european-union>

Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences* 11: 1633-1644.

Verburg P, Soepboer W, Limpiada R, Espaldon M, Sharifa M, Veldkamp A (2002) Modelling the spatial dynamics of regional land use: The CLUE-S model. *Environmental Management* 30: 91-405.

Viaggi D, Signorotti C, Marconi V, Raggi M (2015) Do agri-environmental schemes contribute to high nature value farmland? A case study in Emilia-Romagna (Italy). *Ecological Indicators* 59: 62-69.

#### **f. The address of the project public website, if applicable as well as relevant contact details.**

Project website: [www.envieval.eu](http://www.envieval.eu)

Project coordinator: Dr. Gerald Schwarz.

Contact details: email: [gerald.schwarz@thuenen.de](mailto:gerald.schwarz@thuenen.de)

