ENVIEVAL

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

(Project Reference: 312071)

Area 2.1.4: Socioeconomic research and support to policies

KBBE.2012.1.4-08:

Report D9.5
Methodological Handbook
for the evaluation of environmental impacts of RDPs

Authors: Zymantas Morkvenas (BEF), Kestutis Navickas (BEF), Justas Gulbinas (BEF), Aivaras Jefanovas (BEF), Gerald Schwarz (TI), Anne Wolff (TI), Frank Offermann (TI), Bernhard Osterburg (TI), Inge Aalders (JHI), David Miller (JHI), Jane Morrice (JHI), George Vlahos (AUA), Alexandra Smyrniotopoulou (AUA), Janne Artell (LUKE), Jyrki Aakkula (LUKE), Heini Toikkanen (LUKE), Andrea Povellato (CREA), Davide Longhitano (CREA), Valentina Lasorella (CREA), Katalin Balázs (SZIE), Péter Tóth (SZIE), Csaba Centeri (SZIE), László Podmaniczky (SZIE)

Approved by Work Package Manager of WP9: Zymantas Morkvenas (BEF)

Date: December 2015

Approved by Project Coordinator: Gerald Schwarz (TI)

Date: December 2015

This document was produced under the terms and conditions of Grant Agreement No. 312071 under the European Union's Seventh Framework Programme for research, technological development and demonstration. It does not necessary reflect the view of the European Union and in no way anticipates the Commission's future policy in this area.

Table of Contents

Abbreviations

AEM	Agri-Environment Measure
AIR	Annual Implementation Report
CAP	Common Agricultural Policy
CF	Carbon Footprint
CMEF	Common Monitoring and Evaluation Framework
CMES	Common Monitoring and Evaluation System
DiD	Difference-in-Difference
DREMFIA	Dynamic multi-REgional sector Model for FInnish Agriculture
EEA	European Environment Agency
EC	European Commission
ENRD	European Network for Rural Development
FADN	Farm Accountancy Data Network
FBI	Farmland Bird Index
FSS	Farm Structure Survey
FU	Functional Unit
GHG	GreenHouse Gas
GNB	Gross Nutrient Balance
GWP	Global Warming Potential
HFC	HydroFluoroCarbons
HNV	High Nature Value
IACS	Integrated Administration and Control System
IPCC	Intergovernmental Panel on Climate Change
ISPRA	Institute for Environmental Protection and Research
LCA	Life Cycle Assessment
LFA	Less Favoured Area
LPIS	Land Parcel Identification System
LU	Livestock Unit
LULUCF	Land Use, Land Use Change and Forestry
NBS	Number of farmland Bird Individuals
NIR	National Inventory Report
RDP	Rural Development Programme
SAC	Special Area of Conservation
SAPM	Survey on Agricultural Production Methods
SWOT	Strengths Weaknesses Opportunities Threats
UAA	Utilized Agricultural Area
UNFCCC	United Nations Framework – Convention on Climate Change

1 Introduction

1.1 Requirements of evaluations of RDPs

Evaluation is firmly established as part of rural development policymaking in the European Union. The requirements for evaluation are set out in regulations for the different RDP programming periods.

1.1.1 Ex-post evaluations of the period 2007 – 2013: The old Common Monitoring and Evaluation Framework (CMEF)

The Rural Development Regulation, Regulation (EC) No 1698/2005, set out the legal basis for spending EU funds on agreed aspects of rural development in the period 2007 - 2013, which also stipulated the process to be used (Bradley et al., 2010). For the evaluation of rural development programmes 2007-2013 in the EU, Member States were requested to collect indicators on characteristics, needs, expenditures and results. Experiences from evaluations during the 1990s led to a standardised evaluation framework for the period 2000-2006. Regulation 1750/99 (Article 42(2)) required the Commission, in consultation with the Member States, to define common evaluation questions with associated criteria and indicators for the rural development programmes 2000-2006, in particular concerning the mid-term and ex-post evaluations. The standardised evaluation framework contained common evaluation questions, criteria and indicators for each of the nine chapters of support according to EC No. 1257/99 (European Commission, 1999) as well as cross-cutting questions and questions concerning indicators at the programme level with the aim to make results more comparable and easier to synthesise (Grajewski and Schrader, 2005).

The ex-post evaluation of rural development programmes should be anchored in the rural development policy objectives and Article 81 of Regulation (EC) No 1698/2005 (European Commission, 2005) which defines that the progress, efficiency and effectiveness of rural development programmes 2007-2013 in relation to their objectives shall be measured 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 Framework (CMEF) was drawn up by the European Commission in collaboration with the Member States with the aim of improving programme performance, ensuring programme accountability and allowing the assessment of the achievement of objectives which have been established (Hofmann et al., 2011). The CMEF provided a single framework for monitoring and evaluation of all rural development measures and is the relevant framework for ex-post evaluation carried out in 2016. The ex post evaluation must, as a minimum, cover the legal requirements, which are set out in Regulation (EC) No 1698/2005.

A common set of input, output, result, impact and baseline indicators for the rural development programmes is defined. Input indicators which refer to the budget or other resources are allocated at each level of assistance. Output indicators measure activities in the programme and include, for example, the number of farms receiving a specific form of support. Building on these, result indicators measure the direct and immediate effect of the activities and provide information on changes such as the performance of beneficiaries measured in physical or monetary terms (e.g. number of jobs created). Impact indicators go a step further and measure effects on its beneficiaries both at the level of the intervention but also more generally within the programme area. Environmental impact indicators include reversing biodiversity decline, maintenance of High Nature Value (HNV) farmland and forestry, improvement in water quality and combating climate change. Since common result and impact 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 regions define additional result and impact indicators in a flexible manner, but in accordance with the general principles of the CMEF. A detailed explanation of the legal framework and requirements of the ex-post evaluation of the RDP programming period 2007 – 2013 is provided by the European Evaluation Helpdesk (http://enrd.ec.europa.eu/enrd-static/app_templates/enrd_assets/pdf/evaluation/epe_master.pdf).

1.1.2 Enhanced Annual Implementation Reports (AIRs) and ex-post evaluations in the current programming period 2014 – 2020: The new CMEF and CMES

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) to measure the performance of the whole CAP (both Pillar I - direct payments to farmers and market measures, and Pillar II - rural development measures). More specifically for rural development (Pillar II), there is a Common Monitoring and Evaluation System (CMES), which is part of the CMEF and is set out by:

- the common provisions regulation (Regulation (EU) No 1303/2013), which defines the common monitoring and evaluation elements for the European Structural and Investment Funds (ESI); and
- the rural development regulation (Regulation (EU)No 1305/2013), which addresses the specificities for the rural development programmes.

CMEF – The compilation of rules and procedures necessary for evaluating the whole CAP
CMES - The rules and procedures within the CMEF which relate to rural development (Pillar II of the CAP).

Council Regulation (EC) 1305/2013 on support for rural development by the European Agricultural Fund for Rural Development (EAFRD) obliges all EU Member States to establish a system of ex-ante evaluations, annual implementation reports and ex-post evaluations for each rural development programme (RDP) (Art. 75 to Art. 79) (European Commission, 2013). This regulation specifies the objectives of monitoring and evaluation (Art. 68), the required use of indicators, including the establishment of common indicators (Art. 8, 67, 69), data provision and data management (Art. 69, 70, 71, 76, 78). Regulation 1305/2013 also lays down requirements in relation to monitoring and evaluation reporting in the AIR and to the Monitoring Committee (MC) (Art. 74, 75, 76) including provision of information on the implementation of the evaluation plan (Art. 8, 76) (European Evaluation Helpdesk, 2015).

Rules for practical application of the CMEF, including the CMES, are laid down in European Commission Implementing Regulation (EU) No 834/2014, which sets up indicators allowing the assessment of progress, effectiveness and efficiency of Common Agriculture Policy (Art.1), namely: impact indicators (Section 1 of the Annex) at the level of RD priorities, result indicators (Section 2 of the Annex) at the level of focus areas, output indicators (Section 3 of the Annex) at the level of individual measures and context indicators (Section 4 of the Annex). It further specifies the provision of information by Member States to the Commission for the purpose of monitoring and evaluation (Art.2) (European Evaluation Helpdesk, 2015).

The CMES defines the requirements of monitoring and evaluation, and specifies a number of quantitative common indicators applicable to each programme. 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. This approach is more strategic and consistent than earlier evaluation approaches.

According to the CMES, each of the measures of Pillar 2, is connected to a number of output indicators. Output indicators monitor what has been directly obtained by programme beneficiaries against committed, contracted and implemented funds though RDP measures. Those can be reported through the established CMES and do not require particular evaluation tasks.

The measures are programmed under a set of focus areas. Member States / regions set quantified targets against these focus areas. Result indicators (with some result indicators also being the target indicators) are used to assess to what extent the specific objectives, defined for each of the RDP focus area, have been achieved within the group of programme´s beneficiaries. The assessment of the result indicators forms the basis for answering the focus area related evaluation questions. These focus area-related evaluation questions have to be answered in the 2017 and 2019 enhanced AIRs and in the ex post evaluation in 2024.

Note: To assess to what extent the specific objectives, defined for each of the RDP focus area, have been achieved within the group of programme´s beneficiaries requires a counterfactual assessment against non-beneficiaries to enable quantification of net-effects on beneficiaries.

Table 2. Overview of environmental result indicators, focus areas and focus area related evaluation questions

<table>
<tr><th>Result Indicator</th><th>Focus area</th><th>Focus area related evaluation questions</th></tr>
<tr><td>R6: percentage of forest or other wooded areas under management contracts supporting biodiversity</td><td>4a: Restoring and preserving biodiversity (including in NATURA 2000 areas and areas of High Nature Value farming) and the state of European landscapes</td><td rowspan="2">To what extent have RDPinterventions supported the restoration, preservation and enhancement of biodiversity including in Natura 2000 areas, areas facing natural or other specific constraints and HNV farming, and the state of European landscape?</td></tr>
<tr><td>R7: percentage of agricultural land under management contracts supporting biodiversity and/or landscapes</td><td>4a: Restoring and preserving biodiversity (including in NATURA 2000 areas and areas of High Nature Value farming) and the state of European landscapes</td></tr>
<tr><td>R8: percentage of agricultural land under management contracts to improve water management</td><td>4b: Improving water management</td><td rowspan="2">To what extent have RDP interventions supported the improvement of water management, including fertilizer and pesticide management?</td></tr>
<tr><td>R9: percentage of forestry land under management contracts to improve water management</td><td>4b: Improving water management</td></tr>
<tr><td>R10: percentage of agricultural land under management contracts to improve soil management and/or prevent soil erosion</td><td>4c: Improving soil management</td><td rowspan="2">To what extent have RDP interventions supported the prevention of soil erosion and improvement of soil management?</td></tr>
<tr><td>R11: percentage of forestry land under management contracts to improve soil management and/or prevent soil erosion</td><td>4c: Improving soil management</td></tr>
<tr><td>R12: percentage of irrigated land switching to more efficient irrigation systems</td><td>5a: Increasing efficiency in water use by agriculture</td><td rowspan="2">To what extent have RDP interventions contributed to increasing efficiency in water use by agriculture?</td></tr>
<tr><td>R13: Increase in efficiency of water use in agriculture in RDP supported projects*</td><td>5a: Increasing efficiency in water use by agriculture</td></tr>
</table>

Result Indicator	Focus area	Focus area related evaluation questions
R14: Increase in efficiency of energy use in agriculture and food-processing in RDP supported projects*	5b: Increasing efficiency in energy use in agriculture and food processing	To what extent have RDP interventions contributed to increasing efficiency in energy use in agriculture and food processing?
R15: Renewable energy produced from supported projects*	5c: Facilitating the supply and use of renewable sources of energy, of by products, wastes, residues and other non-food raw material for purposes of the bio economy	To what extent have RDP interventions contributed to the supply and use of renewable sources of energy, of by-products, wastes, residues and other non-food raw material for purposes of the bio-economy?
R16: percentage of LU (Livestock Unit) concerned by investments in live-stock management in view of reducing GHG (Green House Gas) and/or ammonia emissions	5d: Reducing green gas and ammonia emissions from agriculture	To what extent have RDP interventions contributed to reducing GHG and ammonia emissions from agriculture?
R17: percentage of agricultural land under management contracts targeting reduction of GHG and/or ammonia emissions	5d: Reducing green gas and ammonia emissions from agriculture	
R18: Reduced emissions of methane and nitrous oxide*	5d: Reducing green gas and ammonia emissions from agriculture	
R19: Reduced ammonia emissions*	5d: Reducing green gas and ammonia emissions from agriculture	
R20: percentage of agricultural and forest land under management contracts contributing to carbon sequestration or conservation	5e: Fostering carbon conservation and sequestration in agriculture and forestry	To what extent have RDP interventions supported carbon conservation and sequestration in agriculture and forestry?

* Complementary result indicators

Table 3. Overview of environmental target indicators and focus areas

Target Indicator	Focus Area
T8: percentage of forest/other wooded area under management contracts supporting biodiversity	4a: Restoring and preserving biodiversity (including in NATURA 2000 areas and areas of High Nature Value farming) and the state of European landscapes
T9: percentage of agricultural land under management contracts supporting biodiversity and/or landscapes	4a: Restoring and preserving biodiversity (including in NATURA 2000 areas and areas of High Nature Value farming) and the state of European landscapes
T10: percentage of agricultural land under management contracts to improve water management	4b: Improving water management
T11: percentage of forestry land under management contracts to improve water management	4b: Improving water management
T12: percentage of agricultural land under management contracts to improve soil management and/or prevent soil erosion	4c: Improving soil management
T13: percentage of forestry land under management contracts to improve soil management and/or prevent soil erosion	4c: Improving soil management
T14: percentage of irrigated land switching to more efficient irrigation system	5a: Increasing efficiency in water use by agriculture
T15: Total investment for energy efficiency	5b: Increasing efficiency in energy use in agriculture and food processing

Target Indicator	Focus Area
T16: Total investment in renewable energy production	5c: Facilitating the supply and use of renewable sources of energy, of by products, wastes, residues and other non-food raw material for purposes of the bio economy
T17: percentage of LU concerned by investments in live-stock management in view of reducing GHG and/or ammonia emissions	5d: Reducing green gas and ammonia emissions from agriculture
T18: percentage of agricultural land under management contracts targeting reduction of GHG and/or ammonia emissions	5d: Reducing green gas and ammonia emissions from agriculture
T19: percentage of agricultural and forest land under management contracts contributing to carbon sequestration and conservation	5e: Fostering carbon conservation and sequestration in agriculture and forestry

As evident from the comparison of Tables 1 and 3 some result indicators are also target indicators.

Impact indicators provide the means to assess the extent to which the programme has achieved its strategic objectives (EU and national/regional) set up at the level of programme territory. This forms the basis for answering the horizontal evaluation questions. These horizontal evaluation questions related to EU level objectives have to be answered in the 2019 enhanced AIR and in the ex post evaluation in 2024.

Table 4. Overview of environmental impact indicators, strategic CAP objectives and horizontal evaluation questions

<table>
<tr><th>Impact Indicator</th><th>Strategic CAP objective</th><th>Horizontal evaluation question</th></tr>
<tr><td>I7: Emissions from agriculture</td><td rowspan="7">Sustainable management of natural resources and climate action</td><td rowspan="3">To what extent has the RDP contributed to improving the environment and to achieving the EU Biodiversity strategy target of halting
the loss of biodiversity and the degradation of ecosystem services, and to restore them?</td></tr>
<tr><td>I8: Farmland bird index</td></tr>
<tr><td>I9: High nature value (HNV) farming</td></tr>
<tr><td>I10: Water abstraction in agriculture</td><td rowspan="4">To what extent has the RDP contributed to the CAP objective of ensuring sustainable management of natural resources and climate action?</td></tr>
<tr><td>I11: Water quality: Gross nutrient balance and Nitrates in freshwater</td></tr>
<tr><td>I12: Soil organic matter
in arable land</td></tr>
<tr><td>I13: Soil erosion by water</td></tr>
</table>

Details of legal requirements of the evaluations in the programming period for 2014-2020 are provided in guidelines published by the European Evaluation Helpdesk (e.g. guidelines on ‘Establishing and implementing the evaluation plan of 2014-2020 RDPs’, Working Paper on Common Evaluation Questions for Rural Development Programmes 2014 – 2020, and ‘Assessment of RDP results: how to prepare for reporting on evaluation in 2017’.

1.2 Challenges in environmental evaluations of RDPs

Significant methodological challenges exist for the evaluation of RDPs: (i) linkages between different levels of indicators (e.g. result indicators at measure level, impact indicators at programme level); (ii) linkages between impact 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; (v) quantification of net impacts of RDPs at the macro-level and establishing causal-effects relationships; and, (vi) the influence of environmental impacts of rural development measures of site-specific circumstances, which may take a long time to emerge and often depend on a range of other intervening factors.

Existing RDP monitoring systems focus on administrative output and do not provide information on environmental effects. Extensive data and monitoring requirements at different scales are a key challenge for environmental evaluation of rural development programmes. Some measures focus on individual farms and projects, while few measures operate at larger scales such as catchment. Consequently, existing datasets need to reflect different spatial levels, and spatial units are not consistent across different measures and different regional, national and European databases. Availability of, and access to, long term environmental monitoring data for beneficiaries and non-beneficiaries in combination with key secondary databases are critical factors for successful evaluations. Negotiations to obtain data access should start as early as possible in the evaluation process to account for time-consuming activities in the context of different data protection laws.

Recent methodological developments have improved the understanding and capacity of analysing the impacts of farming and forestry on the provision of public goods (e.g. Reinhard et al., 2013; Chabé-Ferret and Subervie, 2013; Michalek et al., 2012). In addition, advances in indicator development and geographic analysis provide new opportunities to address existing key challenges of evaluating environmental impacts of RDPs (e.g. Targetti et al., 2014; Teillard et al., 2012; Concepción et al., 2012). However, a brief review of Tables 1 to 3 is enough to draw the conclusion that a very limited number of indicators can be used for environmental evaluation purposes without substantial elaboration efforts. Many result indicators are unsuitable for robust counterfactual assessments of environmental net-effects of programmes. But data gaps constrain the effectiveness of direct environmental indicators and advanced methods.

The main challenges for evaluations of environmental impacts of RDPs were validated through stakeholder consultations with evaluators and representatives from managing authorities, monitoring organisations across different countries in EU and the European Evaluation Helpdesk. The results of stakeholder consultation and the method reviews highlight the lack of data on non-participants as a key constraint for the application of more advanced evaluation methods. The findings also highlight the need for innovative approaches to design comparison groups in counterfactuals and a better understanding of the linkages between different scales and levels to overcome the challenge to evaluate impacts across different scales and levels. Stakeholders raised the issue that a better understanding of the linkages between different scales and levels is required to overcome the challenge to evaluate impacts across different scales and levels. The need for new indicators in environmental RDP evaluations was highlighted in particular to improve the ability to establish consistent linkages between the impacts of different measures and the overall programme impact. In addition, evaluation methods such as quantitative models should be fit for purpose and better integrate and link different scales and levels of assessment. That also implies that the scales of the data captured and used have to be compatible with those required for the levels of reporting.

To summarise: the following key challenges were identified through stakeholder consultations:

- Lack of environmental monitoring data
  - Existing environmental monitoring data not compatible with RDP uptake
  - Lack of data for non-participants
  - Lack of strategic sampling
- Evaluations across scales and levels
  - Bottleneck: The question of scale and the gap between effects of individual agreements and (potential) impacts at the regional or the national level
  - Limited experiences with upscaling increases uncertainty ('extrapolation of assumptions').
- Counterfactual development and application
  - Area-wide uptake of policy measures – lack of non-participants (control group)
  - To find matching samples at macro level
  - Consideration of other intervening factors in the design of comparison groups
- Complexity of environmental impacts
  - Complexity of environmental public goods and related indicators requires specific evaluation methods to be applied
  - Applying more elaborated methods in the time-frame of evaluations
  - Applying specific methods in combination with existing databases such as FADN
  - Specific expertise required to apply methods

The wide range of complex challenges for evaluations of environmental results and impacts of RDPs highlight that not one solution, or one specific evaluation method, can address all challenges. Different approaches or methods will provide different methodological contributions to reducing the challenges. The selection of a combination of micro and macro level evaluation approaches and methods will depend on the particular objectives, priorities and scope of an evaluation task, availability and access to data and the experiences, skills and preferences of the managing authorities and evaluators. Thus, what is required is a flexible framework providing step-by-step guidance on the design of evaluation approaches and the potential implications of the decisions made at each step for the cost-effectiveness of the evaluation.

1.3 Objectives and purpose of the handbook

The purpose of the ENVIEVAL Handbook is to provide guidance on a methodological framework for use in the evaluation of the environmental impacts of RDP measures and programmes. The Handbook provides flexible guidance to evaluators and managing authorities on a process for designing evaluations of the net effects of RDP measures and is not an 'off the shelf' recipe book. It focuses on the establishment of a suitable counterfactual and the delivery of an assessment of net effects with a requirement for consistency between micro and macro level evaluations.

The Handbook presents a range of possible solutions (e.g. counterfactual scenario setup or choice of appropriate methods for data analysis) with an aim of selecting the most appropriate options for use in evaluations. The output is a basis for deciding on the most consistent and cost-effective approach.

The Handbook is developed around a logic model of the process of assessing the environmental effects of RDP measures. It functions as a methodological framework, guiding the evaluator and managing authority through the design of the evaluation, suggesting different routes depending upon factors such as data availability or type of indicator selected. The step-by-step flow of the logic model(s) helps in the design of a consistent evaluation workflow.

An aim of the Handbook is to help the evaluator to design the ex-ante evaluation in the overall context of the programme, and RDPs and area of interest, in a step-by-step guide for mid-term and ex-post assessments. Competent authorities can use the handbook as support for the overall design of the Terms of References for RDP evaluations.

The ENVIEVAL Handbook summarizes the key characteristics, strengths and weaknesses of evaluation methods recommended, or available, for use in addressing the evaluation of net effects and environmental impacts with respect to a range of public goods. The conceptual description of the different evaluation steps is supported by practical examples of their application and fact sheets of the indicators and methods tested in the ENVIEVAL project which provide information about how to address evaluation challenges, as defined by stakeholders.

This handbook compliments the range of guidance documents dedicated to the evaluation of RDPs, in particular those developed by the European Evaluation Helpdesk for Rural Development.

1.4 How to use the handbook?

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.

1. The Handbook structure

   The Handbook provides:

   (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

2. The in-depth case study examples are presented to illustrate the:

   (i) the context of potential evaluations tasks for which the logic models have been and can be applied
   (ii) micro- or macro-level application of the logic models

3. The in-depth case study examples also highlight:

   (a) type of indicator
   (a) application of a selected method
   (b) data requirements
   (c) strengths and weakness of the method proposed
   (d) recommendations for its application

   Throughout the Handbook specific points, limitations, or key assumptions are highlighted in boxes.

   The Handbook can be entered at different stages of reading:

   (a) the basis and application of indicators and methods
   (b) the step-by-step explanations of the logic model(s)
   (c) the application of the logic models for example evaluations: biodiversity (wildlife), climate stability, landscape quality
   (d) the set of factsheets which set out a suite of indicators and methods for evaluating the impacts of rural development measures on public goods

Be aware that the approaches presented are not 'one size fits all' solutions.

4. Learn more

   If you wish to know more about the ENVIEVAL project, the policy and methodological background, and the case studies please consult the website, www.envieval.eu.

2 Methodological framework for environmental applications of RDPs

2.1 Overview of the logic model for environmental RDP evaluation

The logic models are designed to assist evaluators to design an evaluation and, managing authorities to assess the feasibility of evaluation plans and quality of the results. The logic models are presented with different layers for the design of counterfactuals, and evaluations at micro and macro levels. An overview is presented in Figure 1.

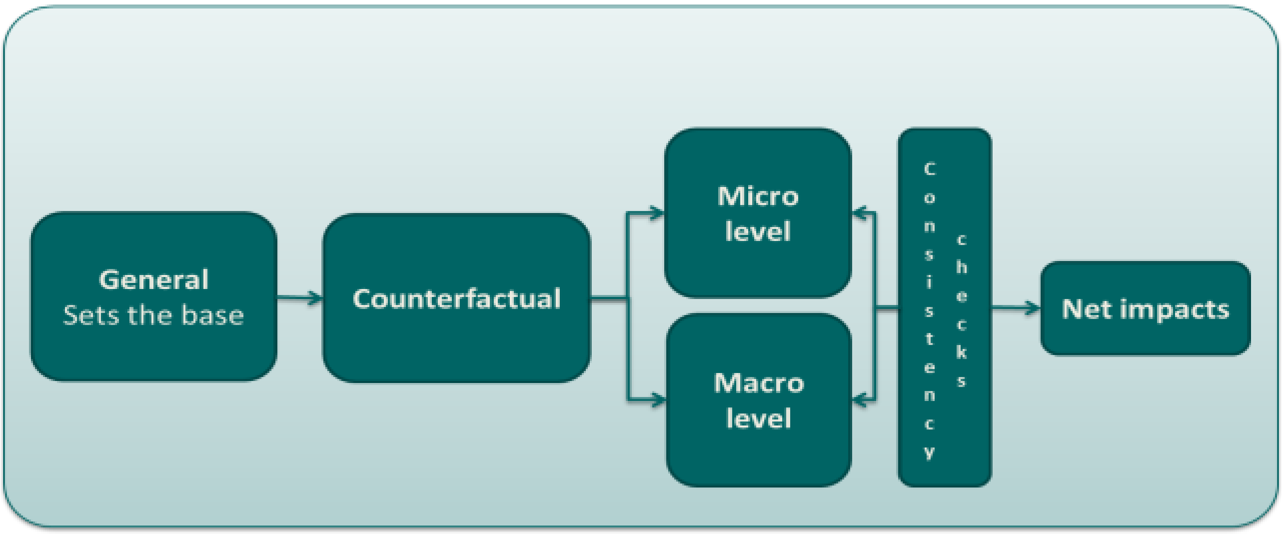

Figure 1: Simplified flowchart of logic models

2.2 Overview of aspects of the cost-effectiveness of evaluation approaches in the application of the logic model

The handbook provides Step-by-Step guidance on the design of cost-effective evaluation approaches for the specific evaluation task to be undertaken. Each Step consists of actions and tests which lead to decisions on the development and implementation of the evaluation task. These decisions affect the cost and effectiveness of the evaluation. This section provides an overview of the main cost-effectiveness aspects of the logic model steps in the context of the evaluation cycle. More detail on the implications for the cost-effectiveness is provided in Sections 3. Figure 2 shows the five phases of an evaluation cycle, the logic model steps and key aspects that influence the cost and effectiveness of the evaluation approach.

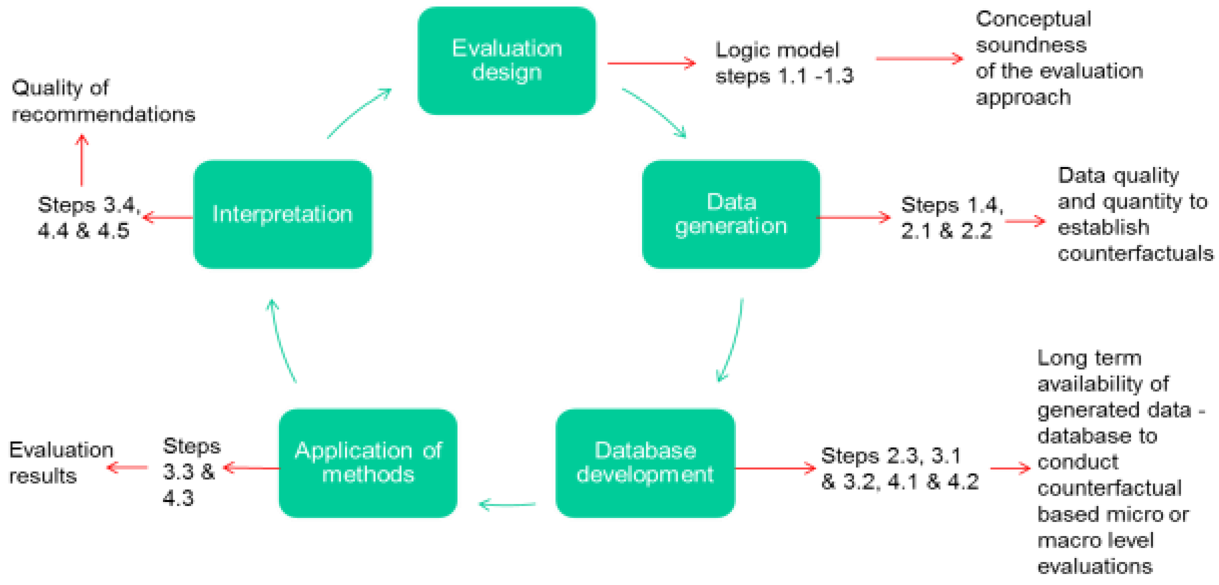

Figure 2: Evaluation cycle, logic model steps and key aspects influencing cost-effectiveness of evaluations

In the first evaluation phase, the evaluation design, several decisions have to be taken, which set the basis for the evaluation approach. In addition to applying the CMES intervention logic (Step 1.1), additional environmental indicators have to be selected if necessary (Step 1.2). This can be associated with high cost as there is an additional work load for the application of the CMES. 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 for the CMES and additional environmental indicators. These activities are crucial for the successful application of the statistical analysis of counterfactuals and have a strong impact on the effectiveness. High cost savings are possible when existing data sources could be discovered and accessed as data collection is usually expensive. Conceptual decisions have further to be drawn on the selection of the unit of analysis (Step 1.3). It can be concluded that decisions at the end of the first evaluation phase are mainly associated with increased labour cost as more time is spent on the development of the evaluation approach. However, importantly, these decisions influence the effectiveness of the evaluation approach at all of its other stages. The right decisions at the beginning of the evaluation process are essential for the successful application of the evaluation method and thus merit 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 (Step 2.1) and the possibilities to construct robust counterfactuals with or without comparison groups with the existing data (Step 2.2) need to be checked. 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 (Steps 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) 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. Consistency checks (Step 3.4 and 4.4) are 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 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.

The following section explains the decisions to be taken at each of the logic model steps in detail followed by examples for the application of the logic model to design evaluation approaches in Section 4.

3 Step-by-step explanation of the logic models to design evaluation approaches

3.1 General: Setting the frame, applying the CMES intervention logic and indicator selection

Step 1.1. Applying the CMES intervention logic

The assessment of environmental impacts of RDPs starts with the requirements and general intervention logic of the CMES, selection of relevant measures, and evaluation questions for environmental objective(s). Selection is then made of output, result and impact indicators which are reviewed in the context of the available data (Figure 3)[1].

Step 1.1 considers if the data available will enable assessments of unintended effects on environment and indirect effects such as deadweight, leverage effects at micro level, and substitution and displacement effects at macro level. Particular attention should be paid to data relating to non-participants.

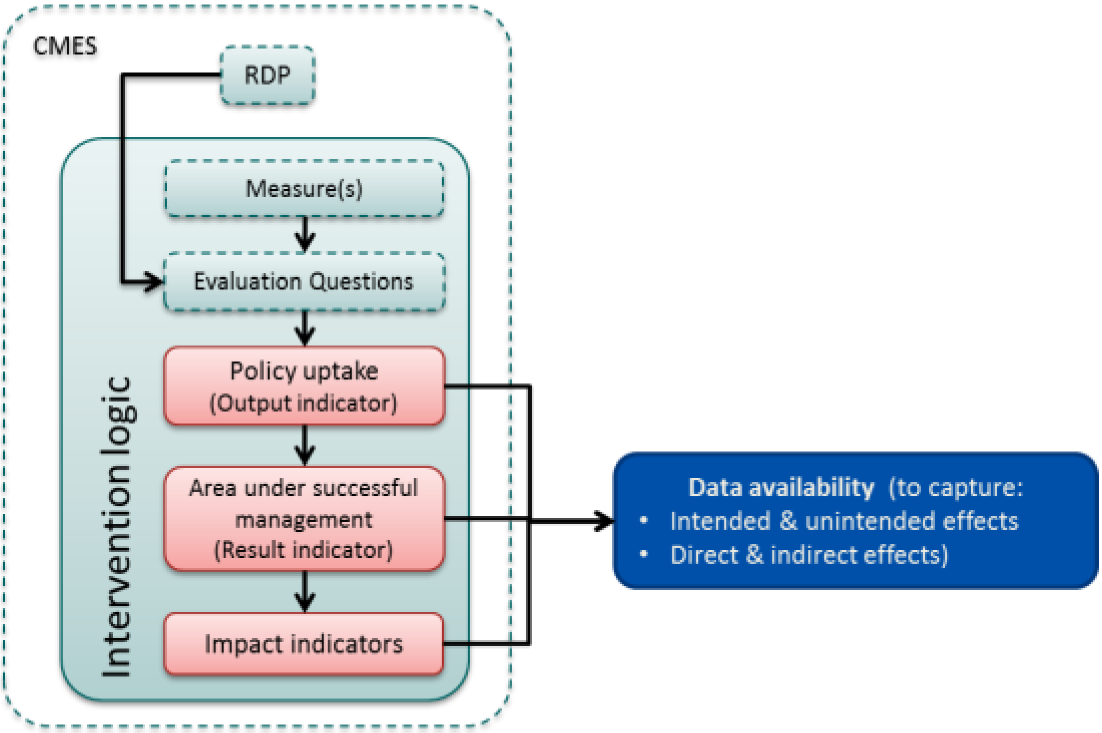

Figure 3: **Step 1.1** CMES requirements

[1] Result indicators of the CMEF relevant to environmental issues are not suitable proxies for measuring environmental changes and, thus for the evaluation of environmental impacts of RD measures and programmes.

Step 1.2. Selecting indicators for public goods

In **Step 1.2** (Figure 4), consideration is given to the environment and public goods for which indicators of impacts of RDP and measures are to be selected and applied. This step is particular important for public goods where the CMES does not provide complementary result and impact indicators.

The CMES provides guidance on general intervention logic. However, the number of environmental impact indicators is limited and it is necessary to:

(i) identify and select indicators suitable for quantifying environmental changes,

(ii) establish robust causal relationships between policy induced land management changes (or livestock management in the case of animal welfare) and the measured environmental change

The suitability of selected indicators should be reviewed in the context of data requirements and the availability of environmental monitoring data.

Fact sheets are available for the following public good indicator tested in addition to existing CMES indicators (to access them, please click on the respective title listed in the cells of the table bellow):

Biodiversity wildlife	**Biodiversity HNV**	**Agricultural landscapes**	**Water quality**	**Animal welfare**
Indicator: Number of farmland bird individuals	Indicator: High Nature Value forestry	Indicator: Shannon Diversity Index	Indicator: Mineral nitrogen content in the soil in autumn (Nmin indicator)	Indicator: Animal-based / result-based indicators: Lameness and mortality rates
Indicator: Number of singing corncrake males	Indicator: High Nature Value farmland	Indicator: Patch shape index	Indicator: Water use/ha	
		Indicator: Land-cover change	Indicator: GNB for the assessment of effects of advisory services	
		Indicator: Visibility of change		
		Indicator: Visual amenity		

As highlighted in **Step 1.1**, it is recommended to pay particular attention to what extent available environmental monitoring data are available for non-participants to enable subsequent consideration of unintended effects on the environment as well as indirect effects such as deadweight effects at micro level and substitution effects at macro level.

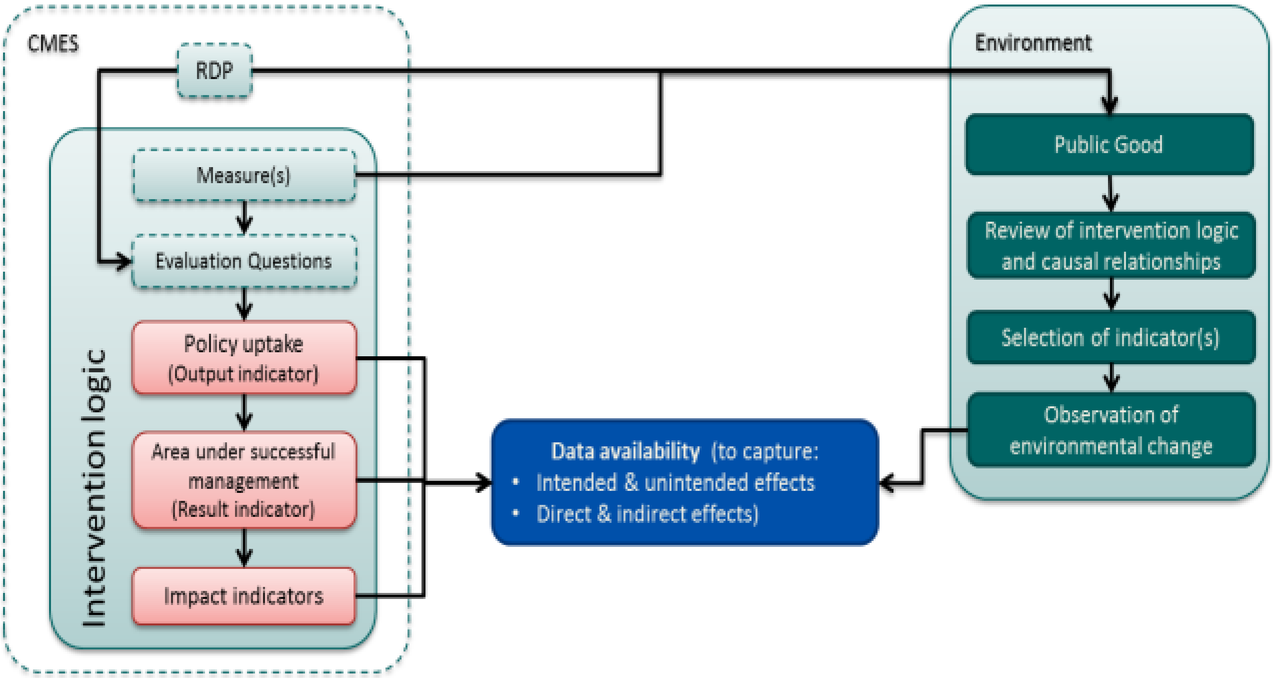

Figure 4: **Step 1.2** Selection of additional environmental indicators

Step 1.3. Definition of unit of analysis

In **Step 1.3**, a common unit of analysis is defined for micro and macro level evaluations (Figure 5). This should be tailored to the public good and environmental objective. The unit of analysis is defined as the "smallest part of an organized system" (parcels, farm as agro-ecosystem, landscape unit, ecological area, sub-catchment area, etc.). It refers to the spatial unit for assessing functional contributions of a system under a specified metric and delimits the analysis and comparison of the system.

The units are characterized by homogeneous activities and addressing scale interdependencies, which is an important aspect for implementation of the logic model. Examples of common units of analysis include parcel and farm (micro), catchment and regional units (macro).

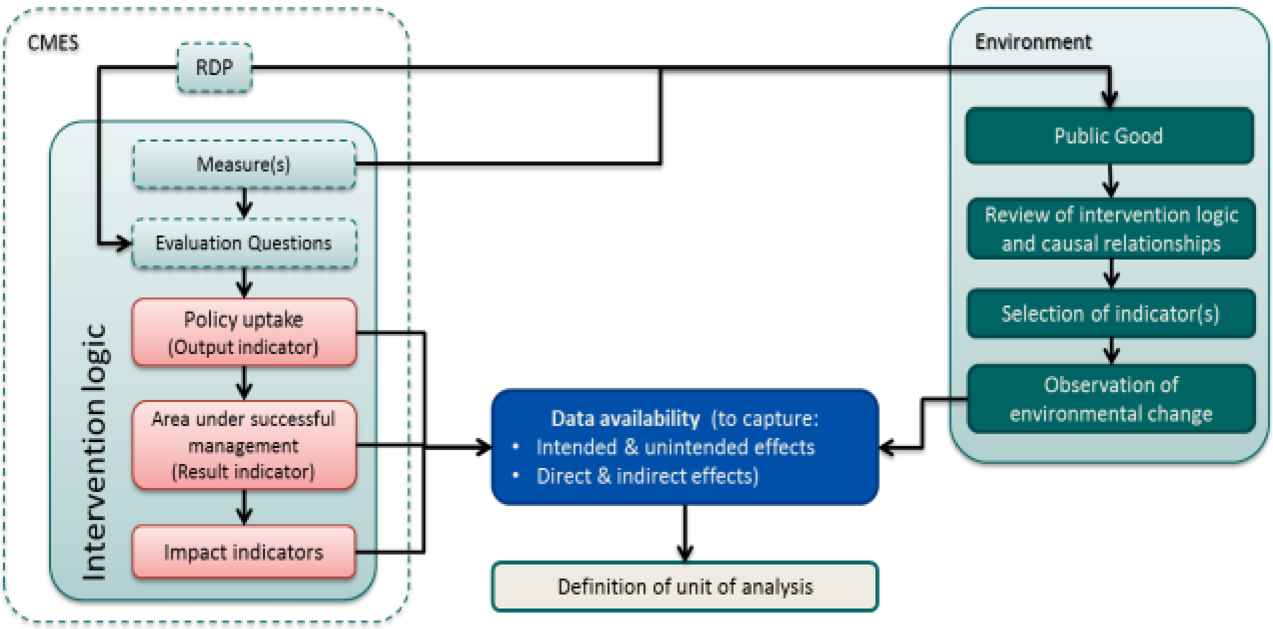

Figure 5: **Step 1.3** Definition of common units of analysis for micro and macro level evaluations

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

In **Step 1.4** (Figure 6), counterfactual based micro level evaluations are designed. The potential requirements are reviewed for the aggregation or upscaling of micro level data and results to a macro level. If that pre-processing is not required, a separate counterfactual-based evaluation design is developed for macro level assessments. In either case, there is a need for consistency checks between micro level and macro levels.

The following section contains an explanation of the steps required for designing and carrying out counterfactual based micro and macro level evaluations of environmental impacts of RDPs.

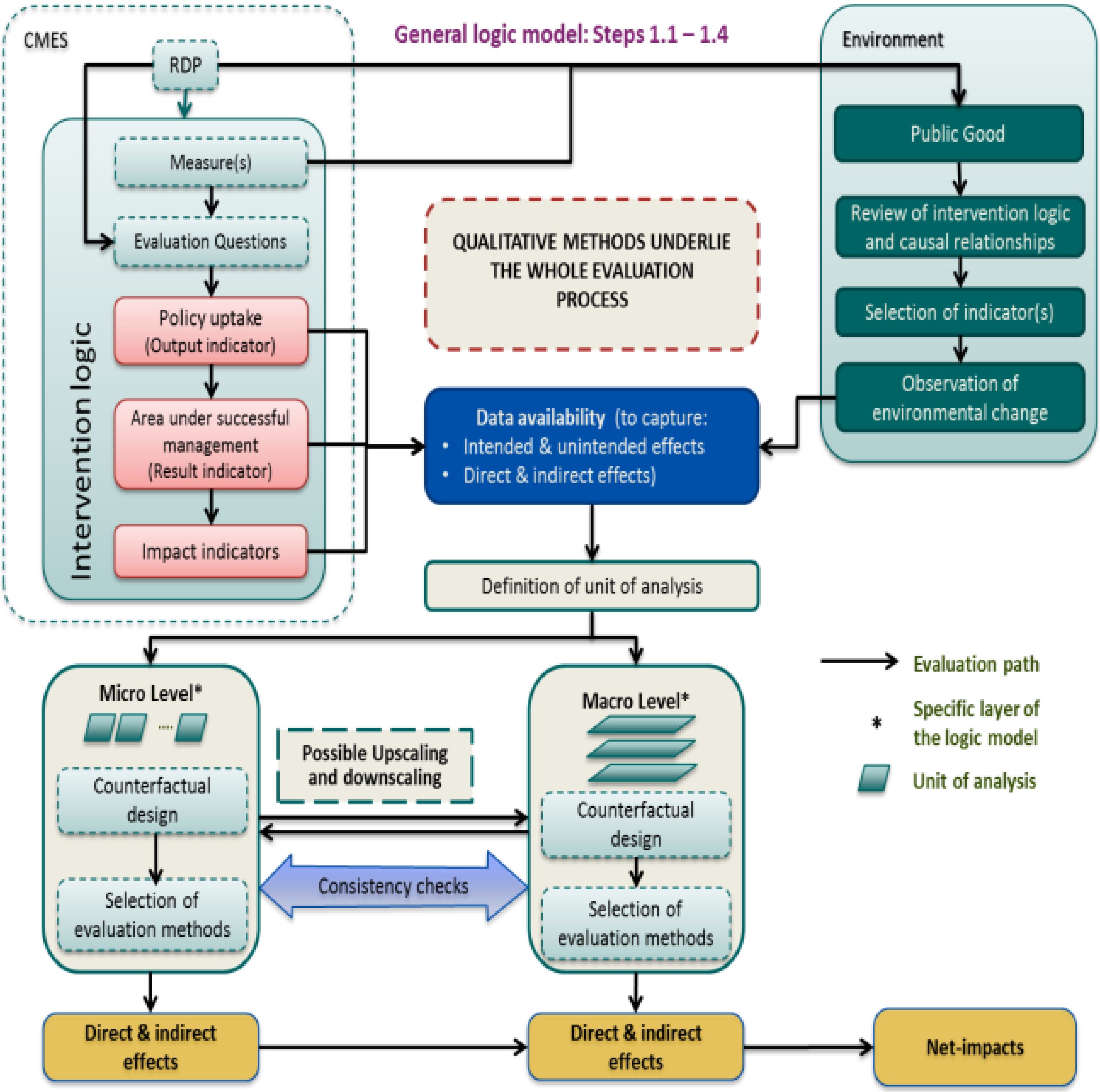

Figure 6: **Step 1.4** Design of counterfactual evaluations at micro and macro levels and assessment of net-impacts

3.2 Designing a Counterfactual

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 stages, providing a set of questions to be answered before deciding on the counterfactual design.

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 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 categorizing the possible methods to design a counterfactual with available data is important even if data is 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. Note that the logic model may convey an impression of preferring quantitative approaches over qualitative. However, careful qualitative assessment should underlie each stage of the logic model[2].

The logic model for designing counterfactuals is shown in Figure 7. In the following text, the logic model is subdivided into component parts and explained. By following the logic model, the evaluator should be able to identify the type(s) of counterfactual they are able to construct with the data available. The logic models at the micro and macro levels are then followed to refine the evaluation options. There are a number of different evaluation methods, or their combinations, available, each with their own specific requirements. We recommend that experts of the identified methods verify the results before embarking on the evaluation process.

With small sample sizes the logic model could lead to a suggestion of a statistical approach that cannot be conducted with the data. Each comparison group in a statistical analysis should, generally, have at least 30 observations at the level of the analysis unit, where the number of observations needed is dependent on the complexity of the evaluation question. Further, we assume that the data are spatially and temporally synchronized with the unit of analysis and the programme period. In cases where this assumption does not hold, the evaluator should very carefully consider the related risks and potential for bias in the evaluation results.

The logic model works best for evaluation cases with a single outcome or result indicator. This is when the intervention logic of a measure or a group of measures is targeting a single environmental outcome. Programme-level evaluations, on the other hand, are often too abstract to be evaluated with a single indicator and, hence, a single evaluation method. Further, the availability of data drives the methods to design a counterfactual measure-by-measure. The logic model provides a structured procedure to assess the range of evaluation options enabling discussion of the impact of different counterfactual scenarios on programme-level evaluation.

[2] The qualitative assessment is important especially in cases where the intervention logic of the measure or policy and data collection have not been well linked in practice.

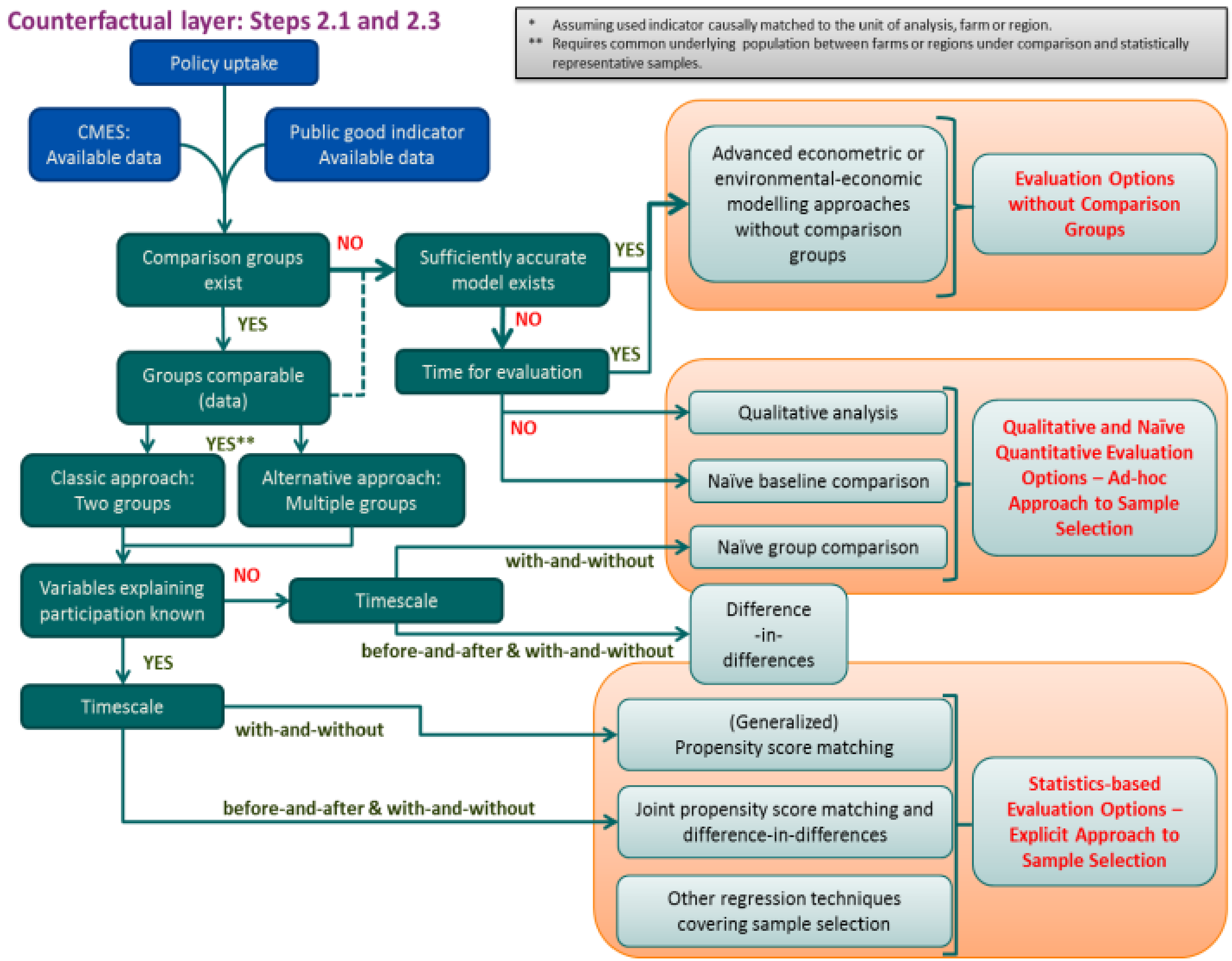

Figure 7: **Steps 2.1 to 2.3** Design of the counterfactual

Step 2.1. Inputs to designing a Counterfactual

In **Step 2.1** (Figure 8) the inputs to the design of the counterfactual are compiled.

Data for the CMES are used to address official, common evaluation question(s) and indicators. Case specific indicators are required when the CMES indicators are not sufficient to measure the environmental impact on public goods at the required level. The number and distribution of the uptake of measures by farms will determine the feasibility of developing comparison groups at later steps of the evaluation.

This Step of the logic model is irrespective of micro or macro level of analysis. Refinements to possible approaches to counterfactuals are undertaken in the micro and macro specific logic models in **Step 3.1** and **Step 4.1**.

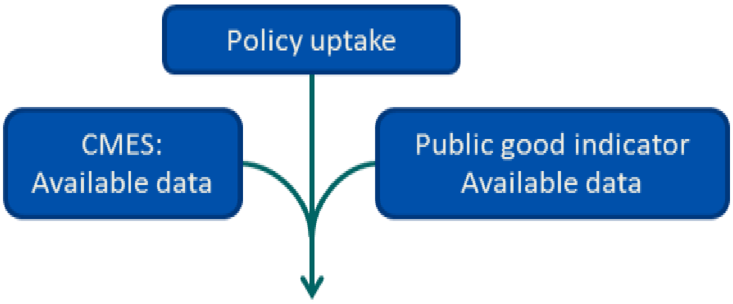

Figure 8: **Step 2.1** Input to the counterfactual logic model

Step 2.2. Defining comparison groups

At **Step 2.2** the data from the input to the counterfactual (**Step 2.1**) are searched to identify the existence of natural comparison groups, such as participants and non-participants. In some cases, every farmer is a participant, or data only exist for participants. Further, the data need to be searched for variables (i.e. characteristics or attributes and external pressures related to the farm or area) that could explain participation in the RDP measure. The latter search helps to consider self-selection to the evaluated programme or measure in the evaluation. The quality and quantity of the data are also assessed to check if they support the construction of comparable groups.

Two approaches to the construction of comparison groups are separated. A Classic Approach uses statistical methods for two comparison groups, while the Alternative Approach uses multiple groups. Methods usable in the Classic Approach can often be accommodated with more than two comparison groups, but may complicate the analysis. The formation of comparison groups (i.e. the first two tests in **Step 2.2**) includes exits to options without comparison groups, should the formation of groups be impossible or infeasible.

Variables that explain participation in the policy, or measure being evaluated, are case-specific and depend on the unit of analysis (e.g. regional versus farm uptake of a measure). These variables should include factors that are targeted in the measure or policy.

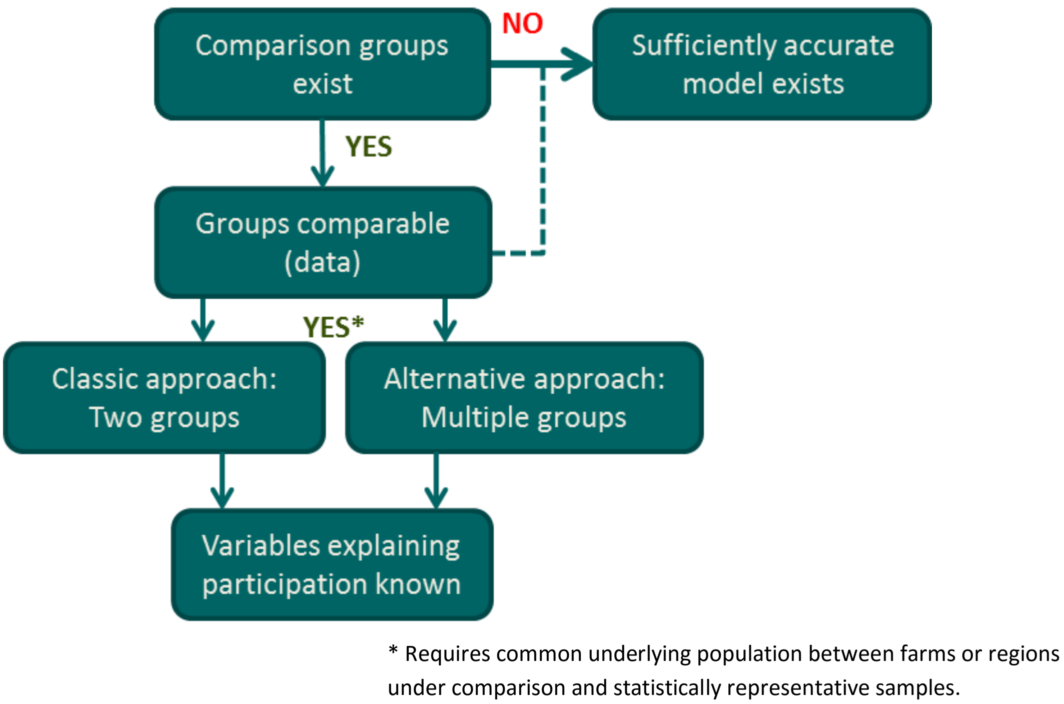

Figure 9: **Step 2.2** Logic model for defining comparison groups

For example, if the policy targets cereal producers below a certain income level in an area of high risk of soil erosion, the minimum requirements for data availability, for all comparison groups are data on:

- farmer income
- production type
- existence of fields in areas of high risk of soils erosion for all comparison groups

However, the meeting the requirements at one observational level (e.g. farm) may be more difficult for another level (e.g. regional).

Table 5 presents commonly expected comparison groups under other intervening factors affecting the end result (e.g. other policy factors). The list reflects the increasing complexity of statistical approaches with multiple effects. Qualitative approaches may be required to describe the potential severity of each effect to limit the number of comparison groups to manageable levels.

Table 5 Determining the number of comparison groups

Participation status in evaluation period	Eligibility rules exist for participation	Internal factors	External factors	Minimum number of groups
Only participants/ non-participants (2)	All eligible (x1)	None (x1)	None (x1)	2
			Historically significant outside pressure at min. one area (+2)	4
		Previous participation status affects	None (x1)	4

Participation status in evaluation period	Eligibility rules exist for participation	Internal factors	External factors	Minimum number of groups
		environmental effects or participation probability (x2)	Historically significant outside pressure at min. one area (+2)	6
	Some non-participants ineligible or in a queue to participate (+1)	None (x1)	None (x1)	3
			Historically significant outside pressure at min. one area (+2)	5
		Previous participation status affects environmental effects or participation probability (x2)	None (x1)	6
			Historically significant outside pressure at min. one area (+2)	8
Participants/ non-participants, also drop outs and/or late joiners (3/4)	All eligible (x1)	None (x1)	None (x1)	3/4
			Historically significant outside pressure at min. one area (+2)	5/6
		Previous participation status affects environmental effects or participation probability (x2)	None (x1)	6/8
			Historically significant outside pressure at min. one area (+2)	8/10
	Some non-participants ineligible or in a queue to participate (+1)	None (x1)	None (x1)	4/5
			Historically significant outside pressure at min. one area (+2)	6/7
		Previous participation status affects environmental effects or participation probability (x2)	None (x1)	8/10
			Historically significant outside pressure at min. one area (+2)	10/12
No non-participants (1)	All eligible (x1)	None (x1)	None (x1)	no statistical comparison possible
		Previous participation status affects environmental effects (x2)	None (x1)	2, note counterfactual is not for inaction
			Historically significant outside pressure at min. one area (+2)	4, note counterfactual is not for inaction
No non-participants but late joiners (2)	All eligible, queues to participate (x1)	None (x1)	None (x1)	2, note counterfactual for partial measure participation
			Historically significant outside pressure at min. one area (+2)	4, note counterfactual for partial measure participation
		Previous participation status affects environmental effects (x2)	None (x1)	4, note counterfactual for partial measure participation
			Historically significant outside pressure at min. one area (+2)	6, note counterfactual for partial measure participation

Step 2.3. Choice of evaluation options

Step 2.3a: Evaluation option without comparison groups

In **Step 2.2** the number of possible comparison groups has been decided.

In the case where no comparison groups can be constructed, **Step 2.3a** follows in which a test (1) is made of whether a sufficiently accurate model exists for explaining participation in the RDP measure (Figure 10: **Step 2.3a** Choice of evaluation options without comparison groups for counterfactual construction. If a suitable model exists that should be used in the evaluation, such as in the Finnish case studies on climate stability and diffuse water pollution (Method: Sector models - DREMFIA model – section 5.1.2).

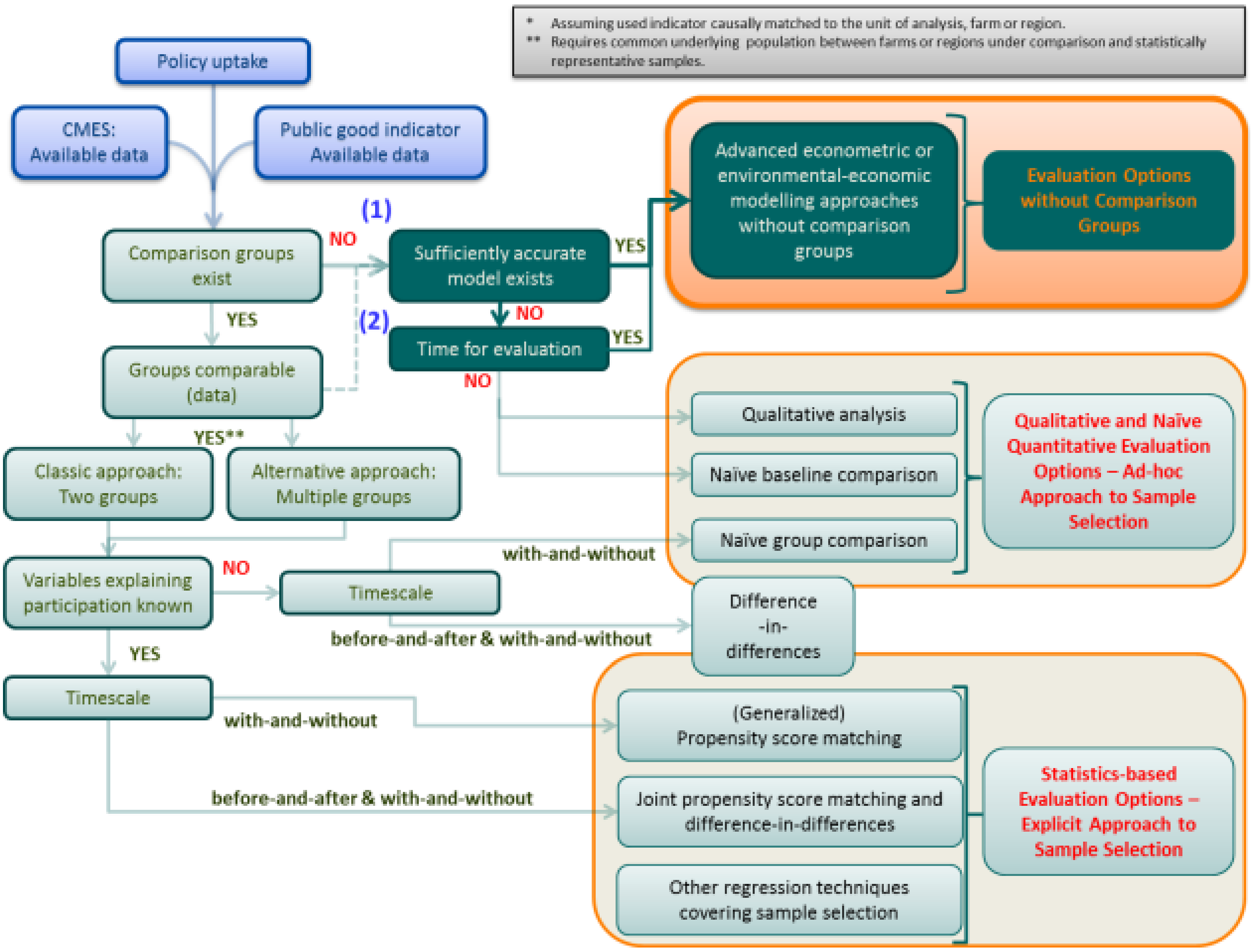

Figure 10: **Step 2.3a** Choice of evaluation options without comparison groups for counterfactual construction.

Existing models have the potential for quick, repeatable, and (possibly) ex-ante evaluations. Sometimes the evaluation question requires a recalibration[3] of the model that can be impossible within the timeframe and resources available for evaluation. Thus, in **Step 2.3a**, if a sufficiently accurate model does not exist,

[3] To accommodate new data, new formats of data or changes in the domain of the model (e.g. eligibility rules, restrictions or market conditions).

and if time and resources allow[4], (2) a model should be developed that enables a consistent counterfactual study of effects in future evaluations.

When using modelled counterfactuals, the evaluator has full control and responsibility in choosing the point of comparison. There are a number of possible scenarios which could be used in modelling, including business-as-usual without the measure being evaluated. The use of scenarios in counterfactuals should support the most credible path should the measure not have been implemented. The evaluator must also decide if other intervening policies or measures should be taken into account in the scenario.

Where there are not sufficient resources to build a model, the evaluator is guided to Step 2.3b.

Step 2.3b: Naïve statistical comparisons or qualitative analysis of counterfactual

Following **Step 2.3a**, in **Step 2.3b** (Figure 11: **Step 2.3b** Choice of naïve statistical comparisons or qualitative analysis of counterfactuals), failing the first two tests (1) and (2) in situations where there is no existing model, or no time to make one, the evaluator can opt for a naïve baseline comparison based on qualitative analysis (6). For example, expert opinions can be used to determine if a trend exists in the development of the environmental indicator. The counterfactual can be based on such a trend and then contrasted to the observed level of the indicator. In this way the evaluator decides the level of environmental impact due to the measure. Therefore the decision, its arguments and how it affects the counterfactual should be explicitly documented with a qualitative sensitivity analysis if possible. Alternatively a fully qualitative approach (7) can be taken to assess the impacts when quantitative data cannot be fully used to support the counterfactual development.

[4] Modelling approaches provide a good evaluation tool especially for widely implemented horizontal policy measures.

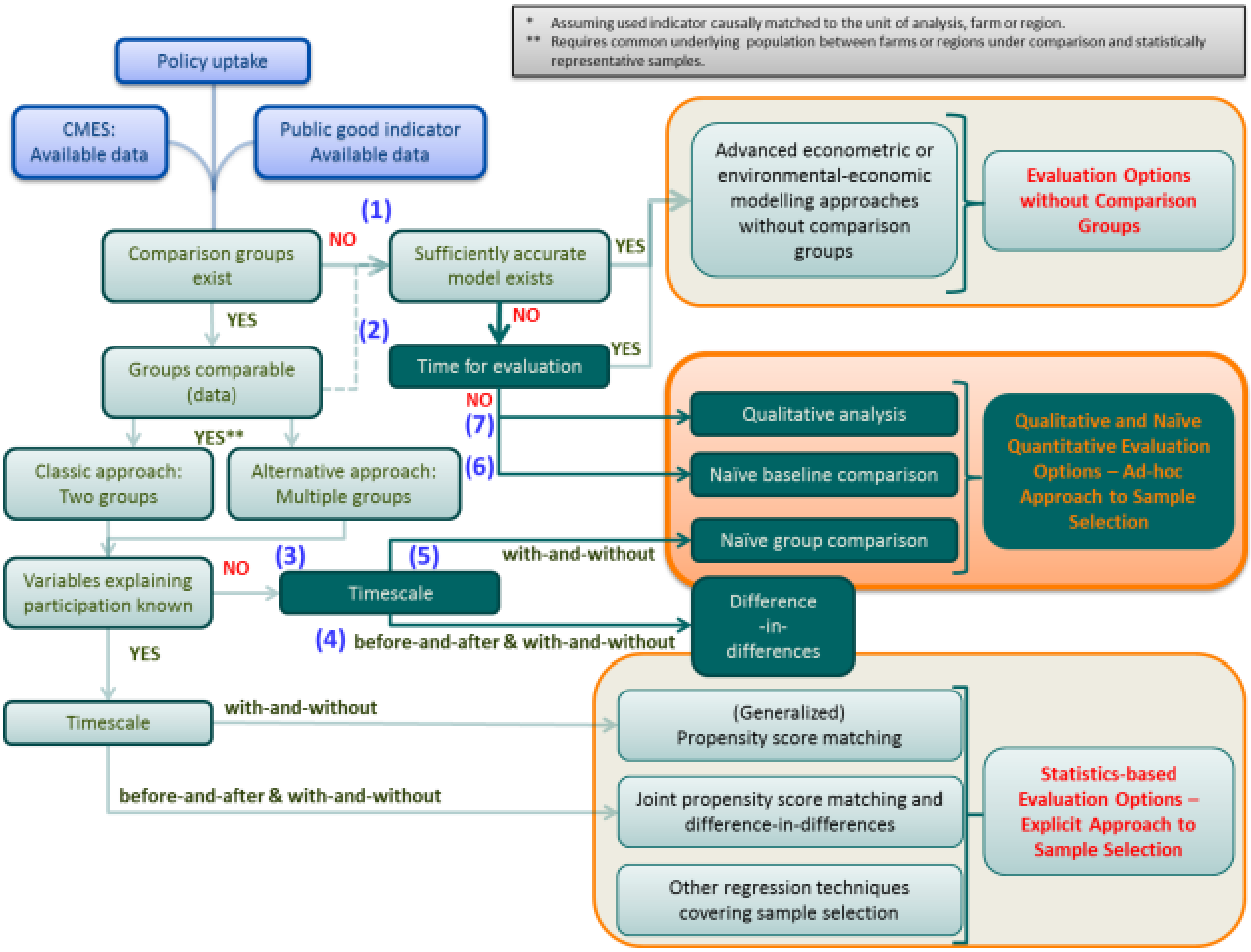

Figure 11: **Step 2.3b** Choice of naïve statistical comparisons or qualitative analysis of counterfactuals

Step 2.3b can also follow **Step 2.2** (3; defining comparison groups), and test of whether data on comparison groups include adequate information that would explain why farmers would participate in, or self-select, the measure being evaluated. If there are no such variables, the test in **Step 2.3b** relates to the timescale of the data, and specifically to determine if the data covers multiple points in time. The definition of the time period can either be based on data which are available only for the end of the evaluation period, or for that available for the beginning and end of the evaluation period[5]. The test is whether there are sufficient data for participants and non-participants, before and after the evaluation period. If yes (4) then the evaluator can use the difference-in-difference (DiD) family of methods.

The DiD can be considered as either a naïve quantitative or a statistics-based approach depending on the underlying assumption on data quality. If the participant and non-participant groups are over different time periods to those of the relevant environmental indicator, the evaluator can use a suitable correction factor for the data.

For example, in the climate stability case study in Italy (see Section 4.2) for details on designing the evaluation approach tested and the fact sheet on the

[5] This will also include intermediate periods if there are late-joiners and drop-outs in programme participation

Method/Indicator: Carbon footprint method in section 5.1.1), the experience and understanding of the data by the evaluator was used when choosing the sample to compare neighbouring participants and non-participants.

If the test on time period of the data identifies that there is only with and without measure (5) for the end of the evaluation period, the choice is a naïve group comparison which uses a quantitative analysis for comparing the environmental indicator across the identified comparison groups. Following such a naïve estimation of the effects, the evaluator should make a qualitative assessment of what internal and external factors are likely to affect the evaluation. For example, if there was likely self-selection of the measure: (i) are the participants similar to non-participants in other respects? (ii) is the bias in the evaluation results likely an over- or underestimate, and what is the magnitude of such bias?

Step 2.3c: Elaborate Statistics-based Evaluation

Step 2.3c (Figure 12: **Step 2.3c** Choice of elaborate statistics-based evaluation for construction of the counterfactual) illustrates the decision links which focus on data on comparison groups which include information that would explain why farmers would participate in, or self-select, the measure being evaluated. These data should include those on farmers who are participating and those who are not participating in the measure for the whole evaluation period. The variables explaining participation in the measure are case specific but typically they include the type of produce, size of production, and the factors that are targeted by the measure.

If data exist for participant and non-participant groups (i.e. with-and-without) and cover the time before-and-after the evaluation period (1) the candidate methods of analysis are the joint propensity score matching and difference-in-differences methods. Such a combination of data is possible where large scale farm monitoring data can be linked with equally precise environmental monitoring data.

If data only exist for the time after the evaluation period (2), the evaluator can use the **propensity score matching**[6] method. But the full potential of this method can only be utilised, if data are available for farms before the programme period for the variables explaining participation, as participation itself may cause these variables to change over the programme period.

The statistical and econometric toolbox provide other regression techniques which can be used to tackle sample selection issues. A summary of these methods is provided in the Review of Counterfactual Methods (Artell, 2013). In several cases, statistical methods are tailored to the specific case, and require the

[6] **Propensity Score Matching** This method compares participant and non-participant farms with a similar propensity to participate in the RDP measure. This alleviates sample selection bias of a naïve group comparison. For this approach to be employed there must be sufficient data for farms which are similar in both groups. If the participants and non-participants have different underlying population, their comparison for impact analysis may not be appropriate using statistical methods.

Generalized Propensity Score Matching This method can accommodate more than two comparison groups, and represent the level of participation by a continuous variable (e.g. different combinations of measures or intensity, different lengths of time in a programme, etc.), but it is more challenging for the evaluator in terms of data requirements and methodological expertise.

evaluator to have appropriate statistical and econometric expertise. To ensure the appropriateness of the analytical approach, qualitative analyses and peer-review processes can be used.

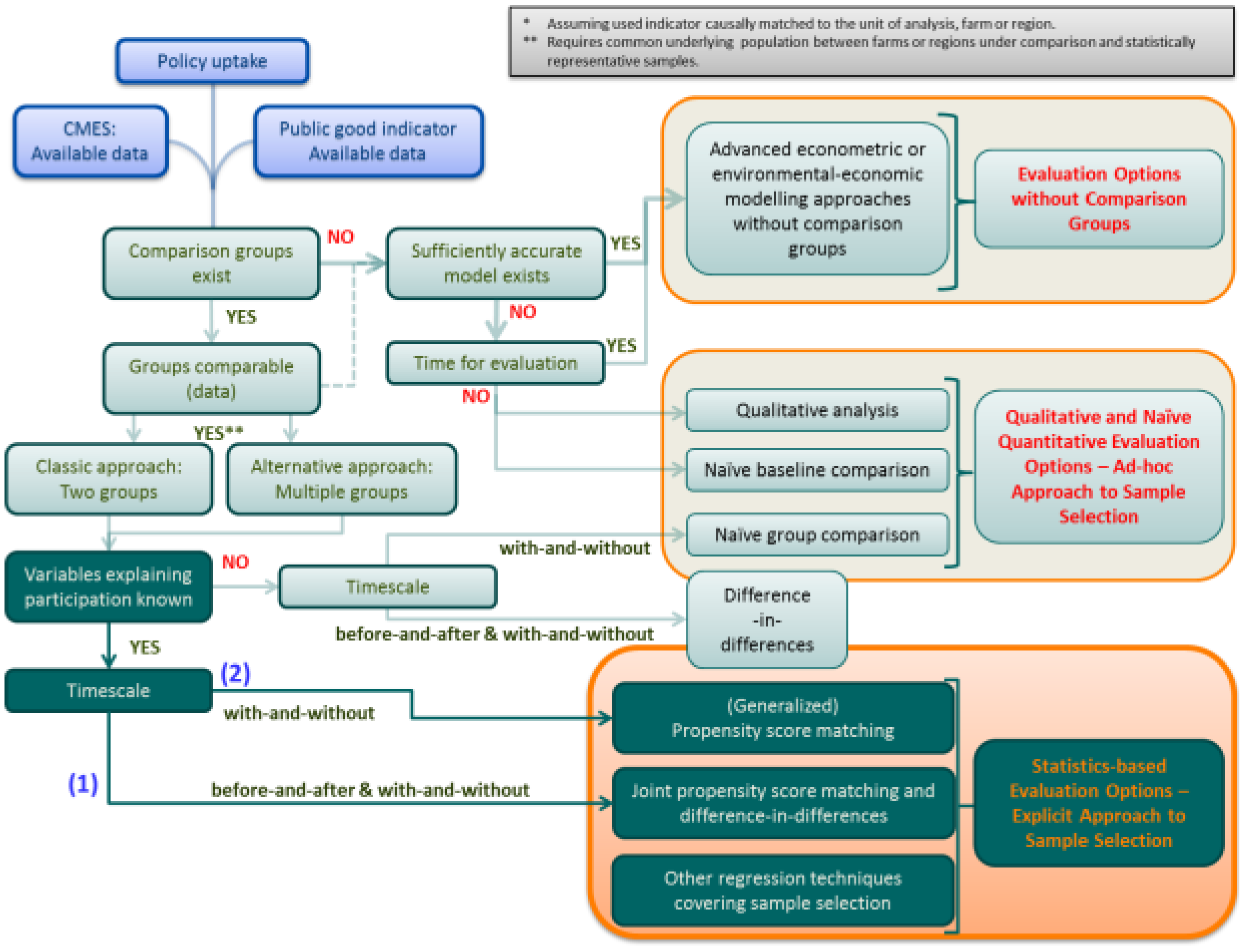

Figure 12: **Step 2.3c** Choice of elaborate statistics-based evaluation for construction of the counterfactual

Collectively, the elaborate statistical methods can be used for any two (or more) comparison groups. If a clear non-participant group data would not be sufficient for statistical analysis a comparison group can be created which is as similar as possible to non-participation in the programme being assessed.

In order to assess synergies between different measures, the use of multiple comparison groups reflecting the uptake of different measures and measure combinations can be a viable option to be explored applying generalised propensity score matching to quantify effects.

The following box summarises key advantages of the three main options.

Step 2.3a: 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.

Step 2.3b: 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 very limited

Step 2.3c: 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.

In cases where comparison groups are available robust statistics-based options dealing with sampling selection issues (Step 2.3c) are preferable in constructing and assessing counterfactuals. But data availability and access might restrict the use of advanced statistics-based options. There is however no need to fall back to simple aggregate comparisons of averages of two groups under Step 2.3b. A number of "second-best" options exist depending on the particular context of the evaluation task at hand:

- Cover sample selection through expert sampling
- Switch methods to less data demanding ones
- Switch from micro to macro-level assessment, enlarge study areas
- Use one point in time instead of before-and-after data
- Compare participant sub-groups
- Use other data
    - alternative indicators
    - 3rd party data, much exists
    - older data with similar conditions
    - alternative data sources for non-participants

3.3 *Assessing environmental impacts at micro level*

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, with tests in the third step for each which leads to different micro-level methods in **Step 3.3**.

Step 3.1. Definition of the unit of analysis and review of selected indicators

Data and indicators have to be selected in order to establish the appropriate links between land management and ecosystem elements and to evaluate the effects of management actions on ecosystems. A clear definition of the unit on analysis allows creating consistent databases supporting statistical comparisons in space and time, if data availability is sufficient. The type of environmental good to be evaluated leads to different approaches in the choice of the observation unit better adapted at measuring environmental outputs.

Farm management surveys and farming system models essentially refer to the farm as the simplest unit of analysis of an agricultural system (micro level), analysed from the point of view of a farmer who decides whether or not to participate in RDPs. By contrast it has to be taken into account that, in some cases, field measurements of environmental outputs could be better referred to different observation units such as farm parcels, monitoring points or landscape units that have to be possibly linked to the farmers' decision to participate or not to RDP.

The logic model for micro-level counterfactuals (**Step 3.1;** Figure 3) starts with a test for data availability for the three counterfactual approaches. Data availability is a significant determinant of the type of unit of analysis that can be used in the evaluation process. These details, combined with those of the requirements for the counterfactual approaches, provide the information required for the selection of suitable indicators to be developed (compare also with tested examples indicators in section 5) and the appropriate observation units.

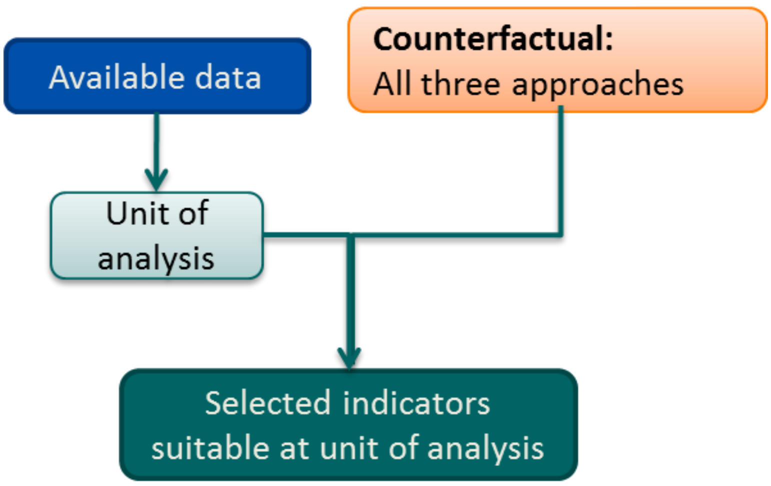

Figure 13: **Step 3.1** Selection of indicators and unit of analysis

Step 3.2. Assessment of data quality

In **Step 3.2**, assessments of the quantity and quality of data are carried out to check if their characteristics are appropriate for the implementation of the methods available for the impact evaluation (Figure 14). Limitations on data quality can often affect the applicability of methods for use in environmental impact assessments leading to a lack of consistent, robust and representative results. Therefore, the logic model includes the identification of potential constraints on the calculation of indicators due to poor data quality. This leads to tests on: (i) the scope for increasing the quantity of data (e.g. number of observations) to assure a better representativeness of the results, and (ii) whether data pre-processing may be required. A test is also made of the availability of spatially-explicit data which may be appropriate for assessments of environmental impacts of certain measures, and be a requirement of some indicators.

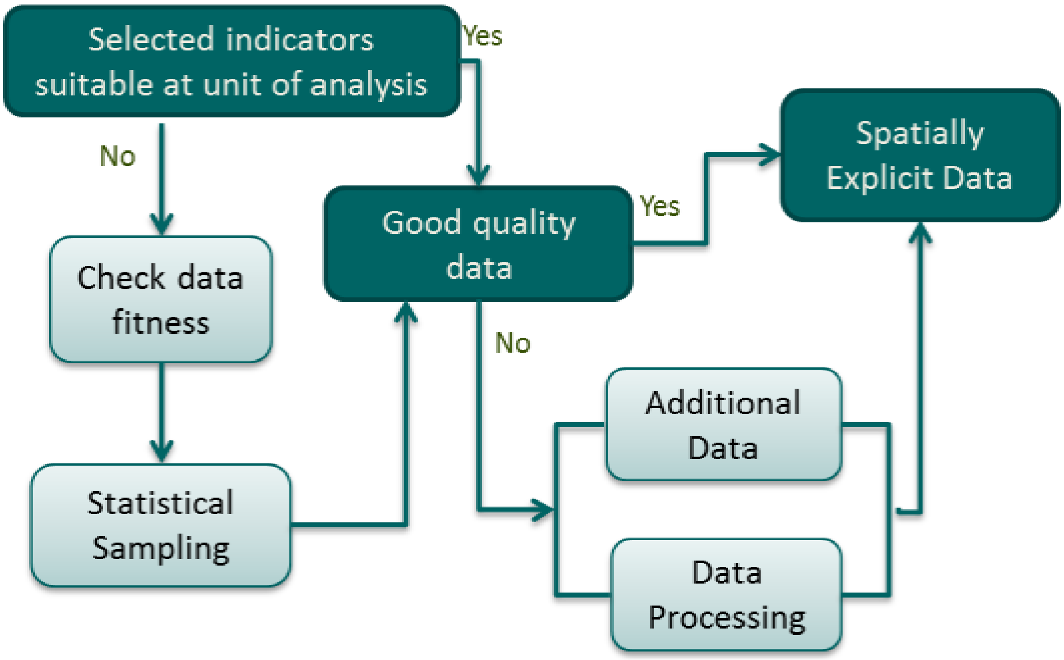

Figure 14: **Step 3.2** Assessment of data quality and quantity

A lack of data can be due to restrictions on access to administrative and statistical databases because of privacy and data protection regulations, or technical issues associated with processes of data collection. The use of qualitative approaches can overcome some problems associated with the lack of data or financial resources for supporting the creation of new databases. They can also be more practical for evaluators to adopt compared to some quantitative methods.

This Step is complete once the relevant data are collated and verified as appropriate for the methods to be used.

Step 3.3. Selection of counterfactual option and micro level method

Step 3.3a: Evaluation Options without Comparison Groups

Without the control group (comparison group) of non-participants, it is not possible to use the statistical approach in the counterfactual analysis (Figure 15).

In these circumstances there are a range of methods available which are based on advanced econometric or environmental-economic modelling approaches and their use should be strongly influenced by the availability of spatially explicit data, and relevant expertise. The principal options which should be considered are:

(i) **Structural models.** Without spatially explicit data, structural models are most appropriate for use at the micro level. Such models use a mathematical approach to studying the link between cause-effect relationships. The method uses a framework for interpreting policy effects due to specific interrelationships amongst endogenous and exogenous variables or factors without the requirements of a comparison group. This allows the capture of effects of specific environmental policies at a micro level. In general, structural models can be used to estimate unobserved or behavioural parameters.

(ii) Integrated models. Where suitable spatial data are available integrated models can be used. These support the evaluation of agri-environment questions in a holistic manner, in particular at the farm scale and its sub-sets, such as areas of cropping areas or land parcels (i.e. the level at which farmers allocate available land and resources to tasks in their production systems). The environmental impacts of changes at the level of the unit of analysis, and associated relationships to environmental impacts can be estimated through the inclusion of links with bio-physical models at a farm scale (e.g. Method: Invest model in section 5.2.1).

(iii) Agent-based modelling (ABM). ABMs enable the coupling of environmental models and relevant social systems and thus the role of social interactions of adaptive, disaggregated (micro-level) human decision-making processes in environmental management. It is possible to identify the role of individuals and an in-depth analysis of forms of organisation (spatial, networks, hierarchies) and interactions between organisational levels. However, given the complexity of ABM models and the intrinsic characteristics of a simulation model they are most appropriately used in ex ante evaluations.

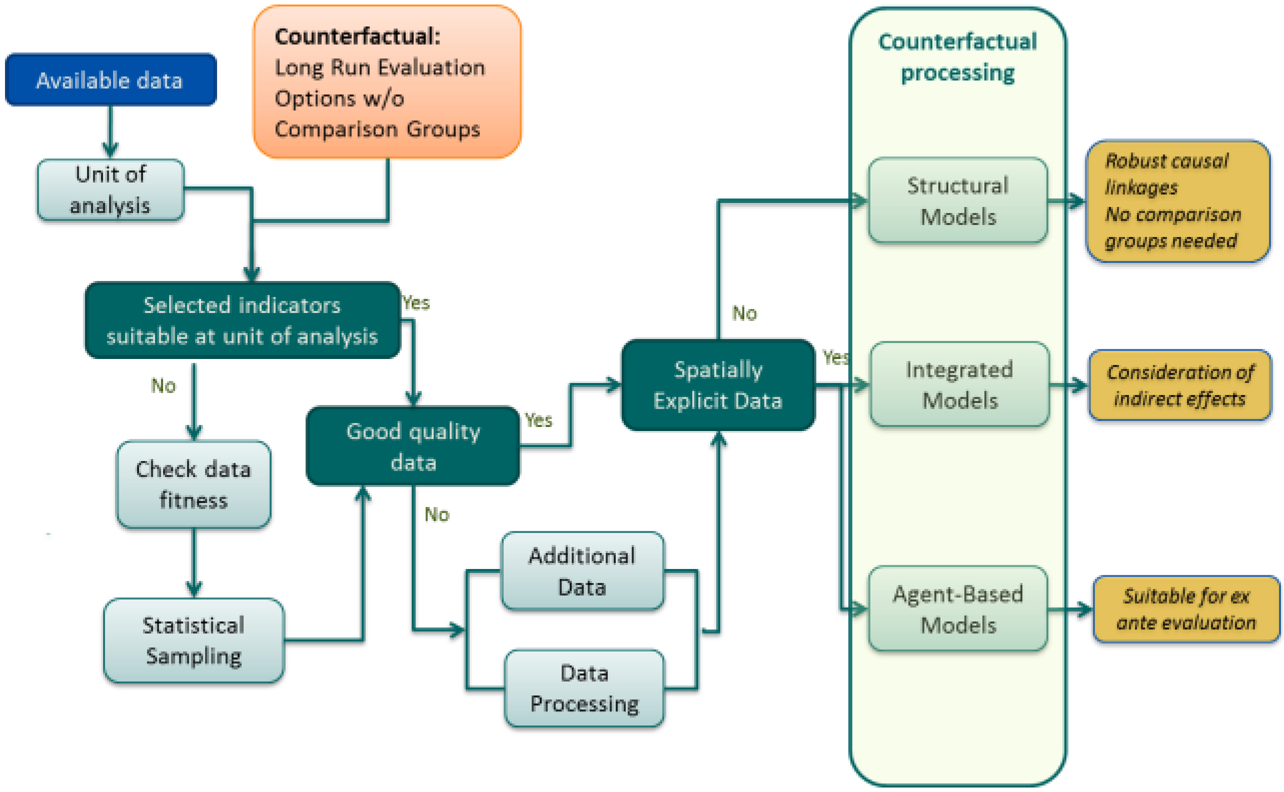

Figure 15: **Step 3.3a** Evaluation Options without comparison groups

Step 3.3b: Qualitative and Naïve Quantitative Evaluation Options

Naïve estimates of counterfactual should be used when data are available on programme participants prior to, and following the programme, but which are not of a sufficient quality and quantity to support the use of elaborate statistics-based approaches to assess net-effects at micro level. The logic model for **Step 3.3b** leads to three approaches (Figure 16):

(i) naïve 'before-after' estimator, which utilises programme data on programme participants to compute programme outcomes for programme participants (without counterfactual);

(ii) naïve 'with-and-without' approach, that use the non-participants as a control group;

(iii) naïve application of a difference-in-difference approach

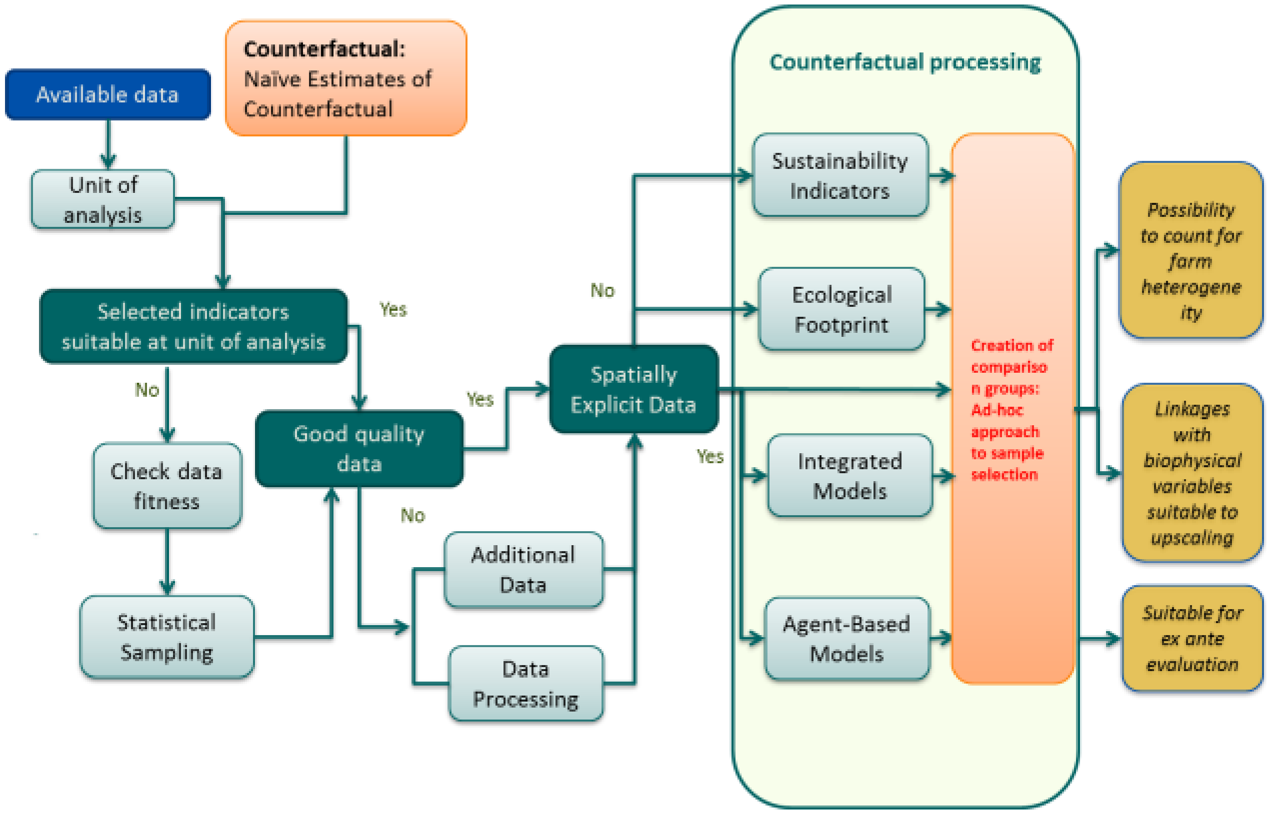

Figure 16: **Step 3.3b** Qualitative and Naïve Quantitative Evaluation Options

These approaches are based on the assumption that, in the absence of the programme, the outcome indicator of participants would be the same as for non-participants. The control group in the naïve comparison of programme participants is represented by the population's average of non-participants. In this approach the data necessary for using a result or impact indicator based on averages in the group of non-participants is usually obtained from various statistical databases (official surveys and/or administrative sources).

In this counterfactual design there may be no need for using a specific method to obtain the information necessary for the assessment, if sufficient self-explanatory variables are available. Otherwise, in **Step 3.3b** options are identified for selecting methods for the creation of the final 'indicator' for naïve and qualitative evaluations, depending upon the availability of suitable spatial data (1). All of the methods identified enable the design of the counterfactual on the basis of data commonly available from official statistical sources available at a local level (e.g. FADN, Census, FSS), and are:

(i) Sustainability indicators. Choice where no spatially-explicit data available.
(ii) Ecological footprint, and composite sustainability indicators. Scope to assess farm heterogeneity due to human environmental actions and thus evaluate policies within a single agricultural system (e.g. Method: Carbon Footprint in section 5.1.1).
(iii) Integrated models. As for Step 3.3a, where spatial data are available.
(iv) Agent-Based Models. As for Step 3.3a, where spatial data are available.

Step 3.3c: Statistics-based Evaluation Options

For Statistics-based evaluations (**Step 3.3c**) four approaches are identified in the logic model, the choice of which is dependent on whether or not there are suitable spatial data available (1):

i) comparing samples of participants and non-participants using matching approaches (e.g. propensity score matching) to test for characteristics and propensity to participate;
ii) conducting an intermediate counterfactual analysis between different participant groups (e.g. participants and late joiners);
iii) using similar but non-eligible farms to represent non-participants (regression discontinuity method);
iv) comparing farms participating and those who are eligible for the program but have not yet been able to join (pipeline method)

For use of an elaborate statistics evaluation it is essential to have sufficient data about general characteristics and the performance of participants and non-participants, and before and after implementation of the RDP. As with the naïve evaluation approach, the application of a specific method for the creation of the final 'indicator' is not necessary if the set of variables used for the statistics-based technique are sufficiently self-explanatory.

The main techniques used to implement this approach are:

(i) **Difference in Differences (DID)**. A difference-in-difference (DID) estimator can be constructed if data on participant and control observations before and after RDP implementation are available. DID compares the effects before and after changes of programme participant, assuming time-invariant unobserved heterogeneity. This method requires the suitable data to be available for two observation periods (i.e. panel data. However DID can be also used on repeated cross-section data, as long as the composition of participant and control groups is properly stable over time..

(ii) **Regression discontinuity design (RDD).** The RDD requires the availability of datasets with variables and observations on eligible and non-eligible units, with a time series of cross-sectional data. The RDD can be used to make an assessment of the effects of programmes that have continuous eligibility.

(iii) **Matching methods such as propensity score matching (PSM).** The matching methods, including PSM, are the most advanced and effective tools of evaluation. They are based on advanced statistical approaches, sufficient data on participants and non-participants, and require the evaluator to have highly quantitative skills. Using this approach at the micro level, the relevant methods identified are the ecological footprint (requiring spatial date), and the integrated models (not requiring spatial data) (Figure 17; 1).

(iv) **Combinations of the above methods.** A commonly suggested combination is the application of PSM in a DiD setting.

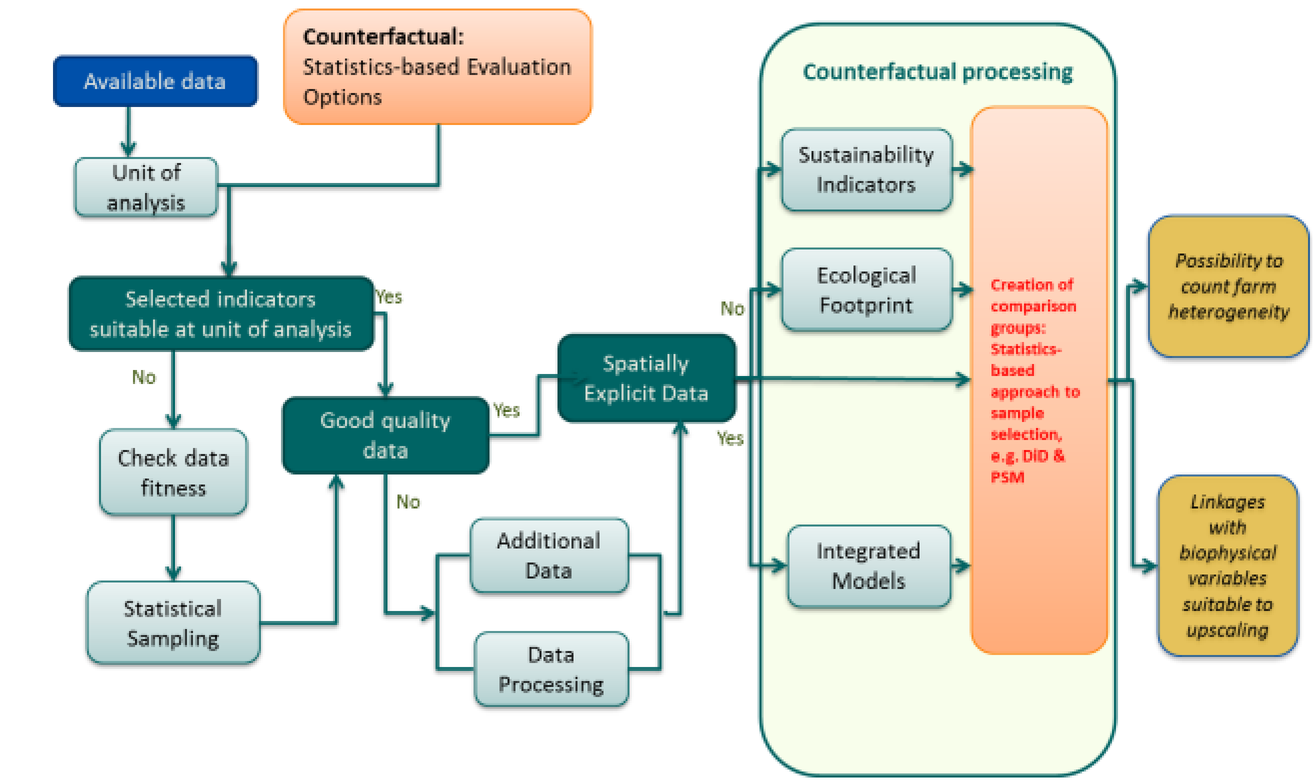

Figure 17: **Step 3.3c** Statistics-based Evaluation Options

Step 3.4. Micro-Macro aggregation and validation

Aggregation and validation of the results is part of the process of analysis (**Step 3.4**, Figure 17).

The aggregation of data from micro to macro levels often has to accommodate multiple data sources, derived from different databases with different metrics and terminology. An evaluation may require use of data at different scales, each with their own most appropriate micro and macro levels which can be a cause of ambiguity. Therefore, the evaluator is directed (1) to a consistency check following, or as part of, the selection of the unit of analysis.

The unit of analysis for the micro level is predominantly that of the farm. Generally, this is the simplest management unit of the agricultural system linked to the implementation of RDP measures. However, 'farm level' can have different meanings in different evaluation exercises so one check required is of consistency in terminology.

On conclusion of the checks for consistency at the unit of analysis tests are made of the selection of indicators, data quality and then the availability and role of spatially explicit data (Step 4.2). At this point in the logic model (2) a check on consistency is also required.

Spatial aggregation from micro to macro-levels consists of up-scaling and aggregating data from a farm level to regional or national levels. However, in the process there can be a loss of detail in the data (e.g. spatial granularity, topological, or in classification), of which the implications for the macro-level evaluation require consideration.

An evaluation of net impacts at a micro level requires:

(i) indirect effects of RDPs to be taken into account, and

(ii) for environmental impacts, the deadweight effects if changes in land use and practice occurred, even without the intervention being evaluated.

Understanding of micro-macro linkages can lead to better definition of indirect effects at a macro level.

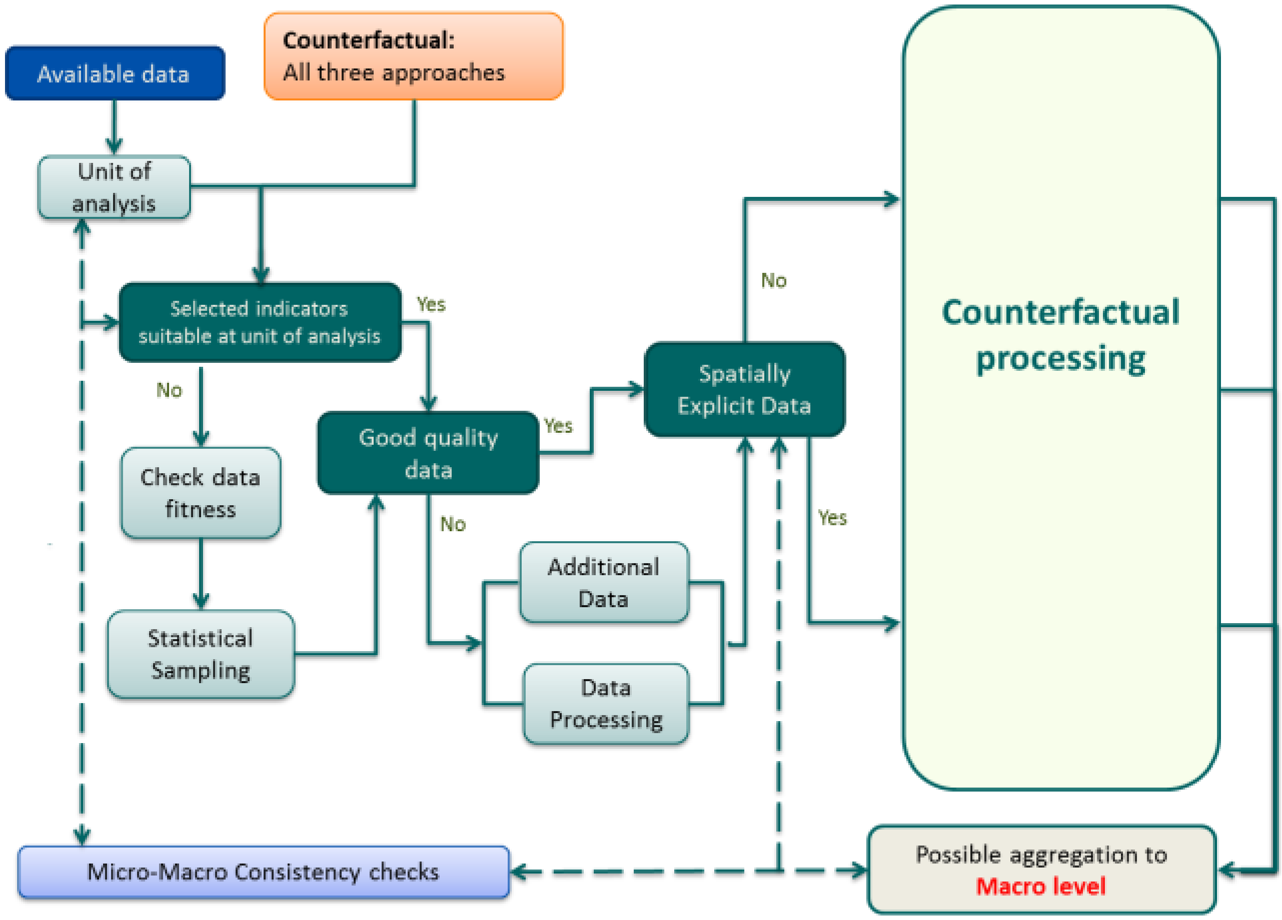

Figure 18: **Step 3.4** Micro-Macro aggregation and validation

3.4 Assessing environmental impacts at macro level

For each of the three possible counterfactual designs identified in **Step 2.3**, an individual macro-level logic model can be created. The first two sub-steps of the workflow of these three logic models are the same. From the third sub-step each counterfactual approach leads to a different set of macro-level methods (See Section 4.3). These have capabilities to assess the impact of RDP on a public good given the unique combination of indicator and data available with which the evaluator is working.

Step 4.1 Definition of the Unit of Analysis and review of selected Indicators

Following selection of the counterfactual approach (**Step 2.3**), the first step in the macro-level logic model (**Step 4.1**) is to identify a unit of analysis (Figure 19). At this stage the evaluator needs to consider the most appropriate unit of analysis given the available data and the counterfactual approach selected. The evaluator also examines, as part of the micro-macro consistency, whether micro-level results can be integrated into the macro-level assessment (i.e. scaled up).

For the unit of analysis selected, the evaluator considers whether there demonstrable consistency between indicators selected for their use at micro and macro levels (compare also with tested examples indicators in section 5).

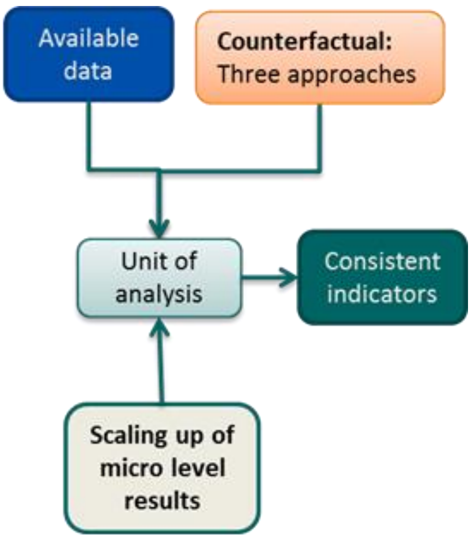

Figure 19: **Step 4.1** Definition of unit of analysis and test of consistency of indicators

Step 4.2 Creation of consistent spatial data

The availability of spatially-explicit data can assist in the provision of quantitative evidence for causal relationships between RPDs and impacts on public goods. The availability of relevant spatial data strongly influences the type of analysis that can be conducted for each of the three counterfactual approaches.

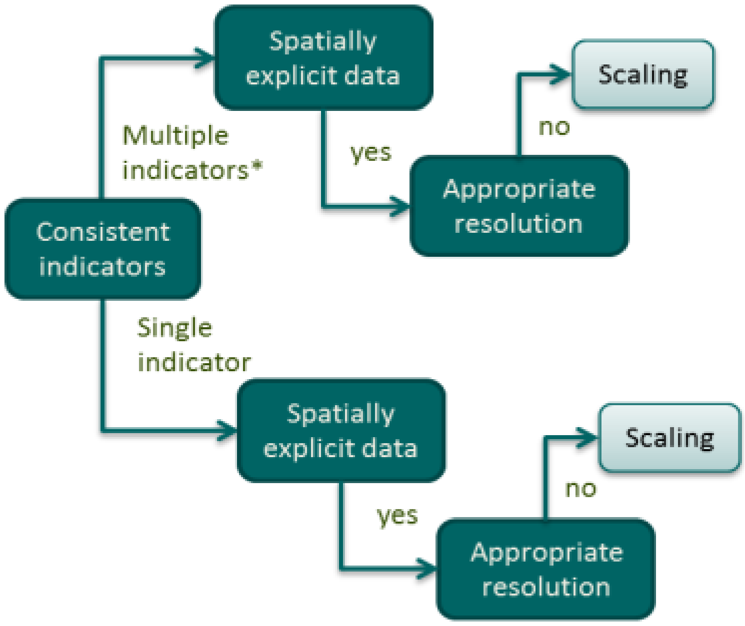

Figure 20: **Step 4.2** Creation of consistent spatial data

In **Step 4.2**, choices are made between approaches which use multiple or single indicators. At a macro level the impact assessment should be able to incorporate multiple indicators as well as single indicators (Figure 20). Methods are available that give the evaluator the option to consider multiple indicators for an impact assessment. However, these differ from the methods for single indicators. Therefore, with the indicators consistent at micro and macro levels the evaluator has to decide whether the assessment will be based on a single indicator or multiple indicators (1).

With multiple indicators, the evaluator will use the macro-level logic model for a single indicator assessment, separately with each indicator, as well as for the multiple indicator assessment. The use of multiple indicator methods for the impact assessment will provide a robust net impact assessment.

Ideally, an assessment will use data of the same spatial resolution. However, if there are suitable spatial data at different spatial resolutions (e.g. spatial support for economic actors at a municipal scale, water quality data at river basin level, soil data for individual soil units) an extra step is required prior to analysis (2 and 3). At this test the spatial resolution is harmonised either through up-scaling or down-scaling methods to a single resolution.

Step 4.3 Selection of counterfactual option and macro level method

In Step 4.3 the choice of counterfactual design leads to a different set of macro-level methods. The logic model for each counterfactual is set out below.

Step 4.3a: Evaluation Options without Comparison Groups

In **Step 4.3a** the methods are divided into those using non-spatial and spatial data in the modelling process (Figure 21).

Non-spatial data. Computable General Equilibrium (CGE) and Partial Equilibrium (PE) modelling frameworks can be used to deal with situations without comparison groups using non-spatial data for single or multiple indicators. The temporal dimensions of environmental impacts are directly incorporated in the dynamic modelling framework, with policy impacts are quantified based on simulations of before and after

RDPs. Modelling frameworks such as CGE and PE provides the flexibility to simulate different counterfactual scenarios, and regional differentiation enables the interpretation of indirect effects at a macro level (e.g. displacement and substitution effects). Care is required with respect to the assumptions applied to inclusion of policy measures in the modelling framework to ensure that the causal relationships of such measures and related land-management changes are theoretically sound.

Spatial data. Three different modelling methods which can support single and multiple indicators using spatially explicit data are suggested. Each of these methods has a unique contribution to make to the evaluation.

(iv) Spatial econometrics can be used to disentangle the external impacts of other intervening factors from the environmental changes which can be directly attributed to the policy measure or programme[7].

(v) Landscape metrics can be used to explore causal linkages and consistency of micro-macro linkages. The use of changes in landscape spatial metrics of factors associated with land cover or use, including their visibility, which are associated with RDP measures. Local environmental characteristics are included, and explicit analysis of micro and macro levels are combined consistently (see Method: Landscape metrics in section 5.4.2).

(vi) Spatial general and partial equilibrium models can be used to explicitly capture spatial and sectoral substitution effects (see Method: Sector models in section 5.1.2).

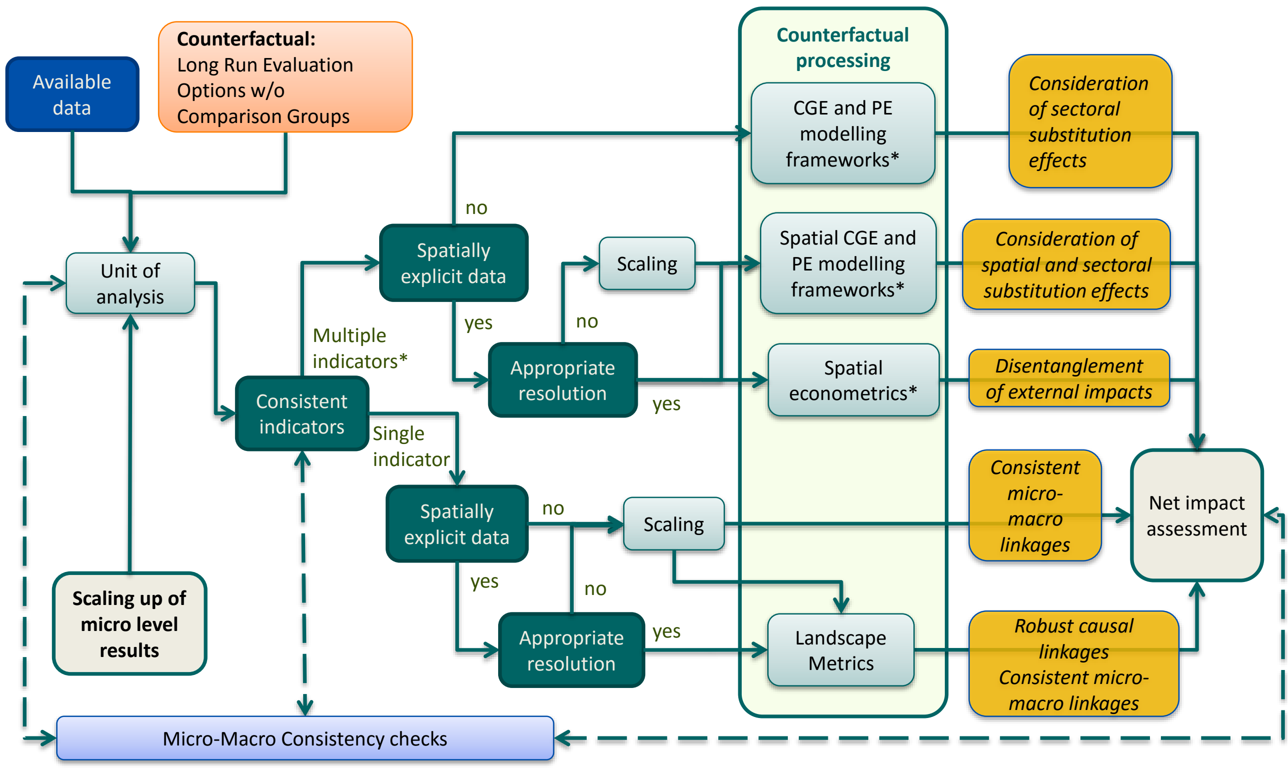

Figure 21: **Step 4.3a** Evaluation option without comparison groups

---

[7] Reinhard and Linderhof (2013) conclude that spatial econometric models provide a suitable methodology to assess environmental impacts of RDPs at a macro level which allows the incorporation of counterfactuals through analysing regions with different spending on the measures and different historic and prospective pathways of development of biodiversity and water quality. However, substantial data requirements for these methods can limit an application for macro-level evaluations of environmental impacts.

Step 4.3b: Qualitative and Naïve Quantitative Evaluation Options

Step **4.3b** presents the logic model where the choice was a Naïve Quantitative Evaluation with an ad hoc approach to sample selection (Figure 22, 1), which leads to a range of methods for consideration by the evaluator:

(i) Ecological foot-printing. In the case of multiple indicators with no spatial data this approach will allow the evaluator to examine non-spatial and spatial macro-level heterogeneity.

(ii) Multi-criteria evaluation. As for ecological foot-printing, this approach also allows the evaluator to examine non-spatial and spatial macro-level heterogeneity.

(iii) Multi-functional zoning is a spatial multi-criteria analysis which can include consideration of the spatially-correlated heterogeneity of multiple indicators.

(iv) Hierarchical sampling. In the case of a single indicator without spatially-explicit data, this method can be used to explore the consistency of micro and macro linkages.

(v) Landscape metrics. For spatially-explicit data, spatial statistics can assist in establishing robust causal linkages and examine consistent micro-macro linkages. For specific public goods such as landscape and HNV, specific spatial statistics such as landscape metrics are established methods of assessing such public goods.

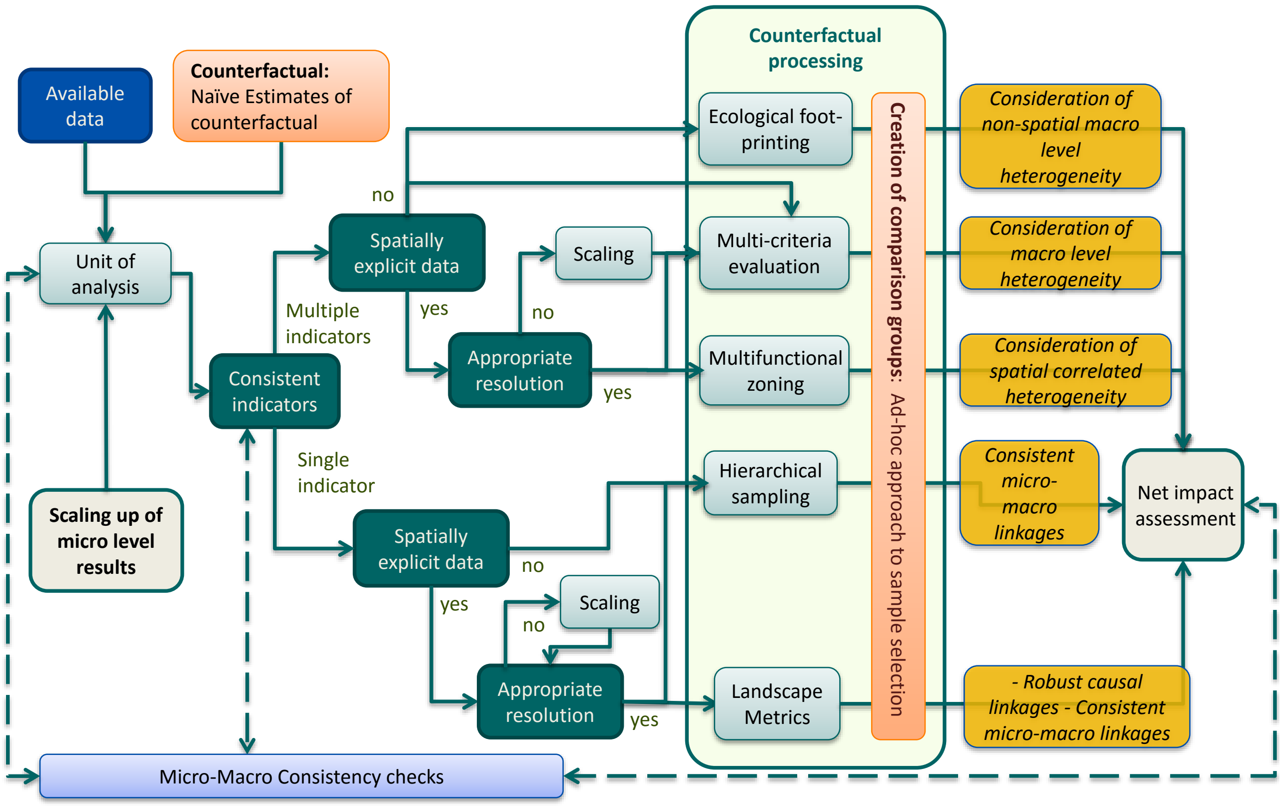

Figure 22: **Step 4.3b** Counterfactual Qualitative and Naïve Quantitative Evaluation Options

Step 4.3c: Statistics-based Evaluation Options

Step 4.3c shows the most comprehensive approach to evaluation, that provided by a Statistics-based Evaluation with an explicit approach to sample selection (Figure 23). The methods used in this approach to developing a counterfactual can make use of single or multiple indicators for the assessment. With access to good quality data, the evaluator will be able to build an impact assessment based upon robust causal links, constituent micro and macro linkages, and to disentangle RDP impacts from external impacts.

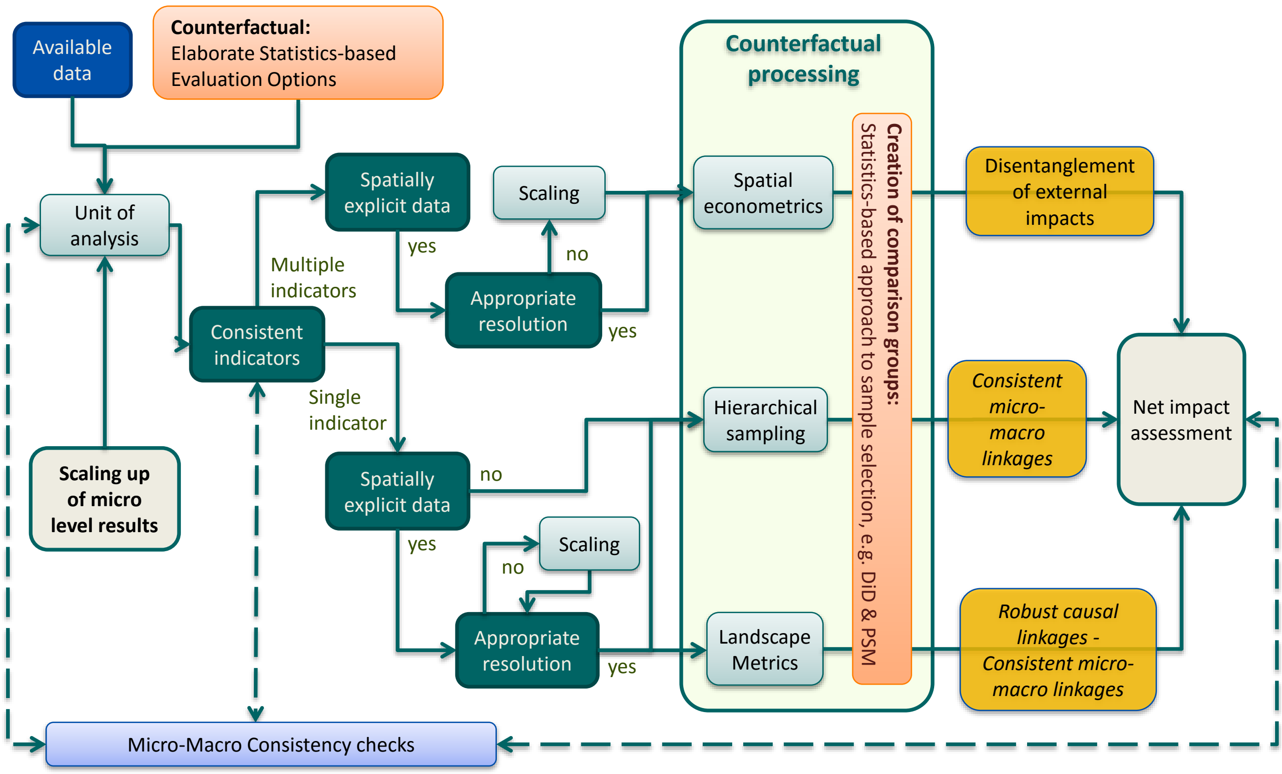

Figure 23: **Step 4.3c** Statistics-based Evaluation Options

The logic model includes tests where the evaluator examines the consistency between the micro and macro levels after selection of the units of analysis (1 – consistency of indicators for micro macro levels) and after the processing of the counterfactual (2 – consistency of micro-macro linkages, and 3 – robust causal linkages, with consistent micro and macro links). These checks on micro-macro consistency ensure that the final results are robust and consistent.

Step 4.4 Net impact assessment

The overarching goal of the evaluation is to assess the net-impacts of the RDPs. The EU evaluation guidance strictly differentiates between: (i) programme outputs (physical units), (ii) results of measures and combinations of measures under the different focus areas, and (iii) programme impacts at macro level on participants and non-participants. The net-effects are those on participants or participating areas as measured with result indicators. Programme impacts reflect medium and long-term effects beyond those immediately felt by participants (or participating areas) which can be observed at community, regional level or programme area. Therefore, these include direct and indirect effects on participants and non-participants, of which the types of indirect effects of relevance at a macro level include substitution, displacement and multiplier effects.

The EU concept of taking into account direct and indirect effects as part of the net-impacts is complex and requires a triangulation of different evaluation methods across micro and macro levels. The main objective of the logic model framework is to identify the specific contributions of methods and approaches and lead the evaluator towards a theoretically sound net-impact assessment.

Examples of particular contributions of selected evaluation methods and approaches at a macro level are highlighted in the yellow boxes in Figure 21, Figure 22 and Figure 23. One of the main contributions of the use of dynamic general or partial equilibrium models (e.g. in assessing climate stability impacts of RDPs) is the explicit consideration of indirect effects such as substitution effects within and between different land-use sectors as well as displacement effects of GHG emissions between different regions. However, consideration of the indirect effects in macro-level assessments requires spatial and non-spatial data of sufficient quality and quantity to enable the application of macro-level modelling or spatial econometric approaches.

Another key contribution to a net-impact assessment is the improvement in the consistency of the results across micro and macro levels. An example presented in Section 5.4.2 is use of landscape metrics to assess landscape or HNV impacts of RDP measures, which includes explicit analysis of micro and macro levels and improves the consistency of the micro-macro linkages of the net-impact assessment.

Macro-level evaluations can build on the upscaling of micro-level results or apply a separate macro level evaluation approach. The latter would include the combination of a 'top-down' macro-level approach (e.g. models with national or regional coverage) for evaluating programme impacts with a 'bottom-up' micro-level approach assessing net-effects of different measures or combinations of measures. Both cases require plausibility and consistency checks to be carried out with respect to the input data, and the comparison of results at different levels. The logic model approach to Net impact assessment and micro-macro consistency is shown in **Step 4.5** (Figure 24).

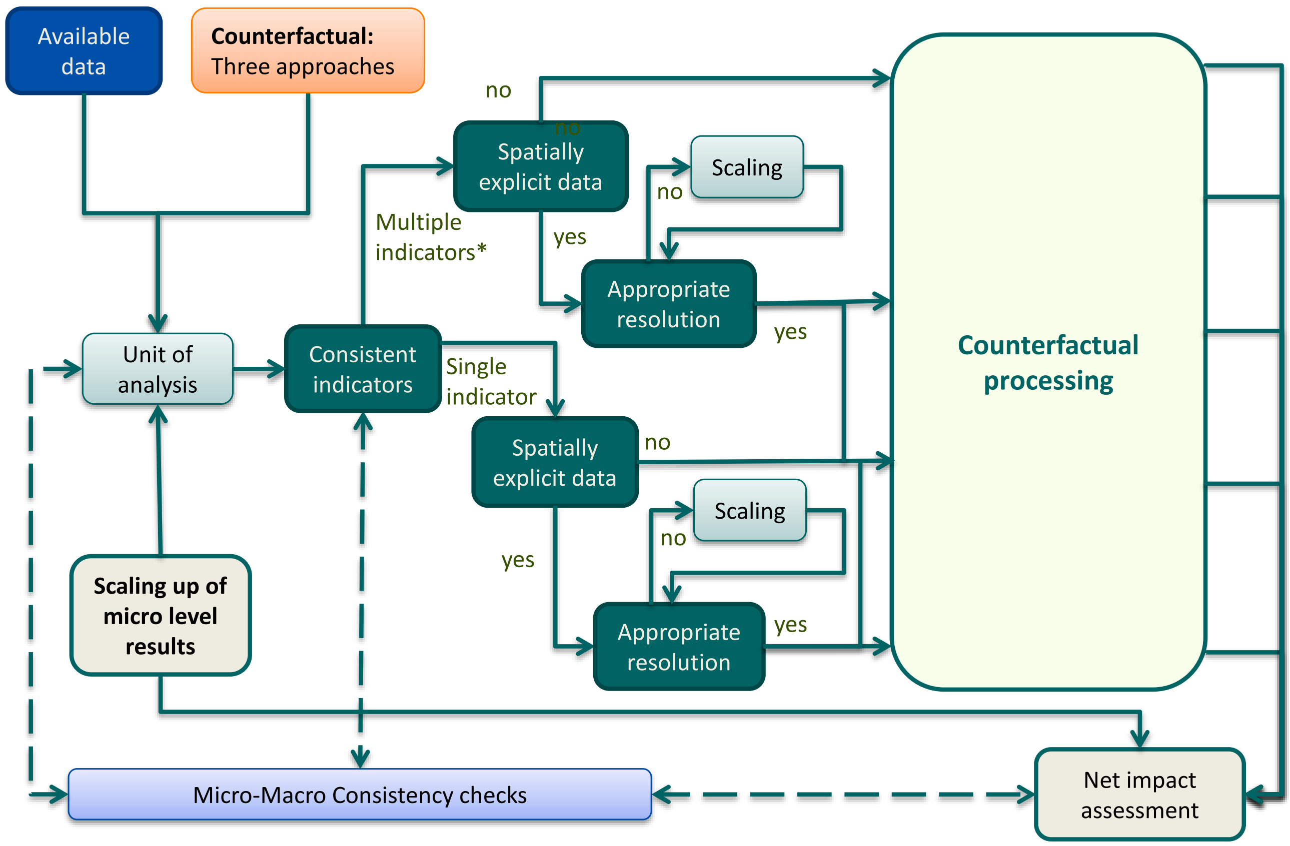

Figure 24: **Steps 4.4 and 4.5** Net impact assessment and micro-macro consistency

Step 4.5 Micro-Macro consistency and validation

A check is required for micro-macro consistency at two points in the evaluation framework:

1. Selection of the unit of analysis and verification of the consistency of the selected indicators (Definition of the Unit of Analysis and review of selected Indicators);

   When considering the unit of analysis for a macro-level assessment, the evaluators takes into account the micro-level assessment and its results. The unit of analysis will have been chosen to be applicable at micro and macro levels. However, the analysis required to change between levels may introduce inconsistencies. For example, a simple aggregation of farm-level (micro level) results to larger areas, possibly along a different scale (e.g. catchments, landscape types, administrative areas), may suffer from inconsistencies in results of the two levels due to farm boundaries not falling within the unit of analysis at the macro level. For example, farms may be in more than one catchment/landscape type, and administrative boundaries may not be consistent with the public good boundaries (e.g. catchment boundaries).

2. Assessment of the net impacts (Net impact assessment);

   For the net impact assessment, it is important that the results of both micro and macro-level assessment are consistent. They are accepted as consistent if results of these assessments show the same trend in relation to impact, even if the evaluator has used different indicators or methods for the assessments.

   The main challenge for the consistency at the micro level is the causality between farm-level action (micro) and beyond farm boundary change. By incorporating consistency checks in the evaluation framework at the beginning and end of the process, evaluators are reminded, at critical moments, to integrate the micro and macro-level assessment which will then benefit the quality of the net impact assessment.

   Environmental impacts consist of direct and indirect effects and are driven by different intervening factors at micro and macro levels. Reviewing the intervention logic and causal relationships between measures, required in land management practices and environmental change needs to achieve clarity on impact pathways across scales, levels and actors. This is an important conceptual basis for assessing the plausibility of the results at micro and macro levels. It may only be possible using qualitative approaches, e.g. where there are data gaps, in which there are reviews of the significance and expected direction of change that indirect effects or other intervening factors might have at micro and macro levels. A key question to be addressed by the evaluator is whether the conceptual basis of the indicators or the framework can be used to explain the differences between micro and macro-level results and thus aid in validating their plausibility.

   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.

4 Examples for the application of the logic models

4.1 Application of logic models for evaluation focussed on Biodiversity Wildlife public good in Hungary

The good availability of biodiversity monitoring data (Common Bird Monitoring Programme, run by BirdLife Hungary at macro level, the baseline data for micro level) enables the case study to address the evaluation challenge of establishing robust causal linkages at both levels between changes in biodiversity indicators and the uptake of relevant rural development measures.

Description of case study region

In the macro-level case study, the total area of Hungary is ca. 93,000 km^2 and the area in the country under agricultural use is 53,000 km^2.

The micro-level case study was carried out in Northern-East Hungary, in Heves-Plain. Heves-Plain is a 610 km^2 area with significant importance for nature conservation, supporting bird species such as the great bustard (Otis tarda) and imperial eagle (Aquila heliaca) or the red-footed falcon (Falco vespertinus). Due to its importance for nature, it has been included in former rural development programmes, for example it has been a High Nature Value Area since 2002, and it includes some Natura 2000 sites, specifically four Special Areas of Conservation (SAC) and one Special Protection Area (SPA). Most of the area is under intensive agricultural use, with significant mosaics of natural and semi-natural grasslands.

Most relevant policy measures

The shortlist of relevant measures was defined based on the estimated causal linkages between the measure and soil/biodiversity issues; the uptake of the measure; estimated data availability and the possibility for detailed evaluation of impact. Based on this, the following measures were selected: - Measure 214 Agri-environmental payments

Figure 25: Map of contracted land parcels under AE measures in Hungary (source: Institute for Geodesy Cartography and Remote Sensing)

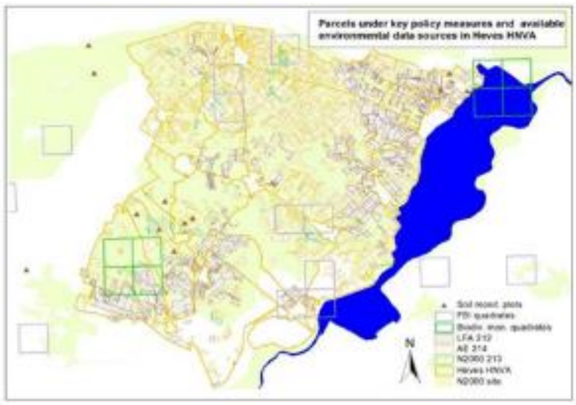

Figure 26: Map of contracted land parcels under AE measures in Heves HNVA (micro level)

During the selection procedure, the 'Natura 2000 and WFD payments on agricultural areas' measure was also taken into consideration but the approaches finally focused on analysing only the AE measures in both micro and macro levels. Natura 2000 payments are available only for grassland areas in Hungary, and the land-use requirements of these areas are not directly linked to the measures but the legal background of Natura 2000 network (269/2007 governmental regulation on land use requirements on Natura 2000 grasslands).

Application of the logic models

Approach: Farmland Bird Index (FBI) and Species numbers of Farmland Birds
Steps of the general layer: CMEF/CMES, selection of environmental indicators, functional units, database concept and data sources

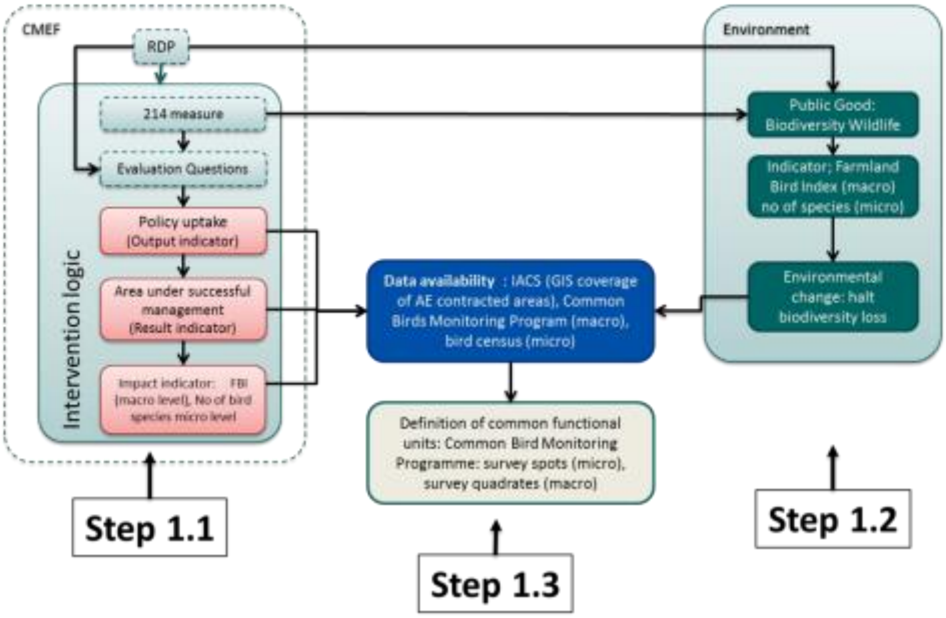

Figure 27: **Step 1.1, 1.2 and 1.3** of the general logic model

Step 1.1. following the CMEF/CMES intervention logic

The case study explored the impact of agri-environmental measures on biodiversity wildlife. The CMEF/CMES defines the Farmland Bird Index impact indicator for evaluating the impacts of the RD measures on the changes in biodiversity as “*the indicator is a composite index that measures the rate of change in the abundance of common bird species at selected sites, i.e. their relative abundance*”. These species, chosen from a list of selected common species at EU level, are dependent on farmland for feeding and nesting, and are not able to thrive in other habitats. The list contains the maximum number of species, from which individual countries select the relevant species. No rare species are included. Population trends are derived from the counts of individual bird species at census sites and modelled as such through time.

Step 1.2. selecting additional environmental indicators

In addition to the Farmland Bird Index, the number of farmland bird species was used at micro level to estimate the micro-level impacts. The indicator of the number of farmland bird species was developed based on the baseline data of the FBI indicator. The indicator was expected to be more sensitive to micro-level impacts, as the functional unit is linked to one distinct parcel of contracted or non-contracted areas. The temporal changes of the number of species may show the real impact of the agri-environmental payments at micro level.

Step 1.3. selecting common functional units for micro and macro-level evaluations

During the design of the appropriate functional units, several attributes will be taken into consideration. Functional units need to be responsive enough for the changes of the particular indicator (Farmland Bird Index, number of farmland bird species), while at the same time being valid for the evaluation level (macro, micro). Besides having clear links to the public good indicator and the level of evaluation, the design of the functional units must be carried out based on the uptake data of the key policy measures (e.g. spatial coverage of the contracted parcels). For the macro-level case study, the baseline data collection quadrats were selected as functional units for the Farmland Bird Index. The Common Birds Monitoring Programme is based on the data collected from 2.5x2.5 km squares. Data collection from these squares started in 1999, and more than 1000 squares have been surveyed. During the evaluation, results from continuously surveyed squares were used.

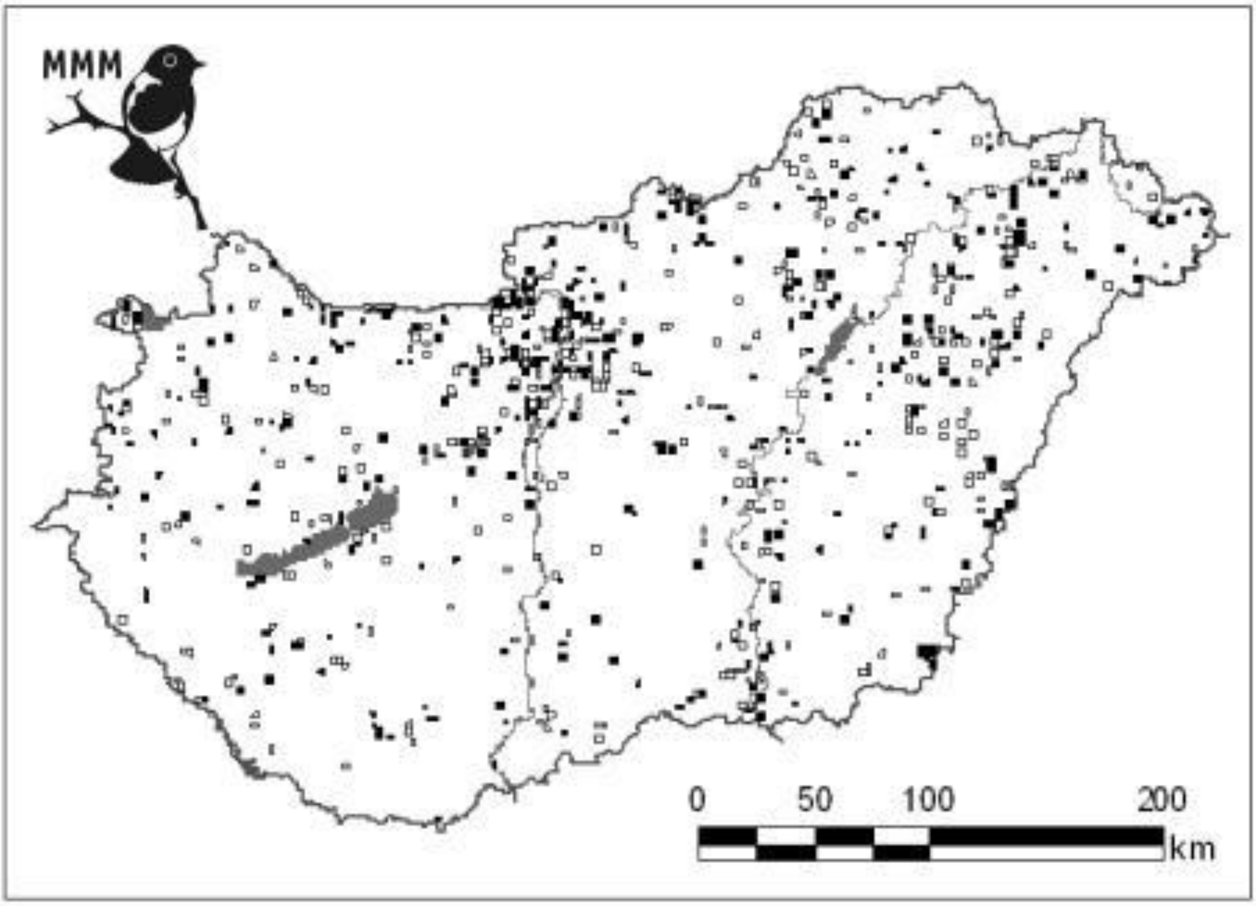

Figure 28: Survey squares of Common Birds Monitoring Program as functional units at macro level (Source: BirdLife Hungary)

At micro level the functional units have been designed to serve both the responsiveness to micro-level impacts, comparability to macro-level results and the further potential development of a micro-level survey. The basic survey spots of the Common Birds Monitoring Programme were selected as functional units. Approximately 300 survey points were taken between 2009 and 2014 from the selected case study area (Hevesi-sík High Nature Value Area) .

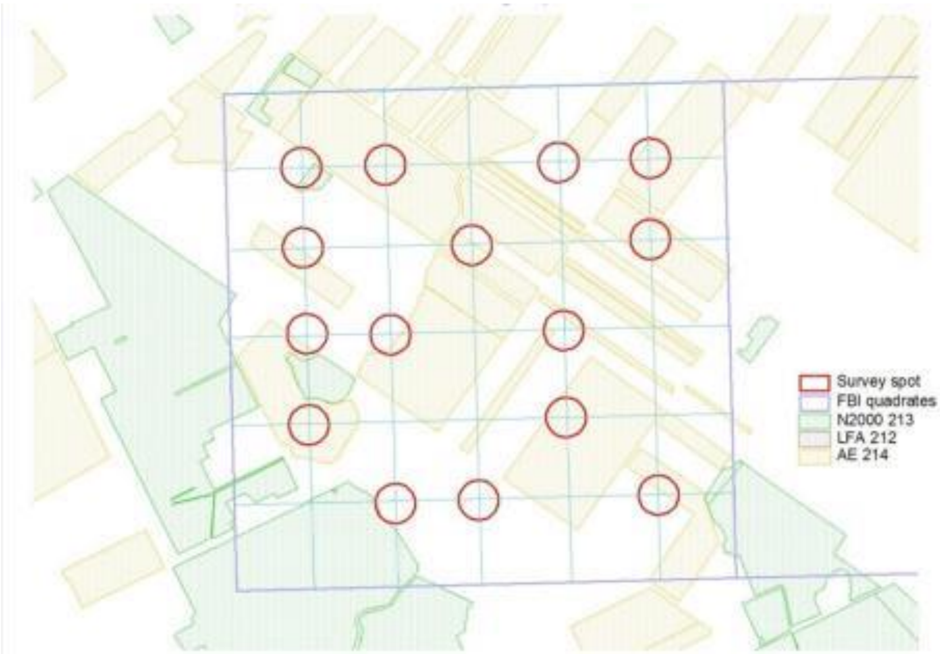

Figure 29: Survey spots of Common Birds Monitoring Programme as functional units at micro level

Data sources

At macro level the data was collected by volunteer surveyors of the BirdLife Hungary and the database development was done by the Monitoring Centre of BirdLife Hungary. The set of bird species was adapted to Hungarian circumstances, with the following species: Falco tinnunculus, Perdix perdix, Coturnix coturnix, Vanellus vanellus, Merops apiaster, Galerida cristata, Alauda arvensis, Anthus campestris, Motacilla flava, Locustella naevia, Sylvia nisoria, Sylvia communis, Lanius collurio, Lanius minor, Sturnus vulgaris, Miliaria calandra (Szép et al. 2012).

At micro level the data was collected by the staff of the Bükki National Park Directorate and local volunteers. The set of bird species was selected following the recommendation of European Bird Census Council for the Central and Eastern Europe region for the possible further replicability in the region. The role of the volunteers during data collection must be highlighted, as without their efforts the appropriate data for the biodiversity wildlife public goods would be inadequate for evaluation purposes.

Relevant spatial coverage of the contracted land parcels of the 214 Agri-environmental measures was provided by the paying agency in GIS format.

Corine Land Cover 1:50.000 was used for the evaluation of background land-use information during the counterfactual design.

Step 1.4. Conceptual decisions on counterfactual micro and / or macro level evaluations

During the design of counterfactual method, a comparison was applied of participants and non-participants before and after measure implementation both in micro and macro level. The time period of the macro-level survey was 1999-2014, while the micro-level case study focused on the previous period of the RD measures (2009-2014).

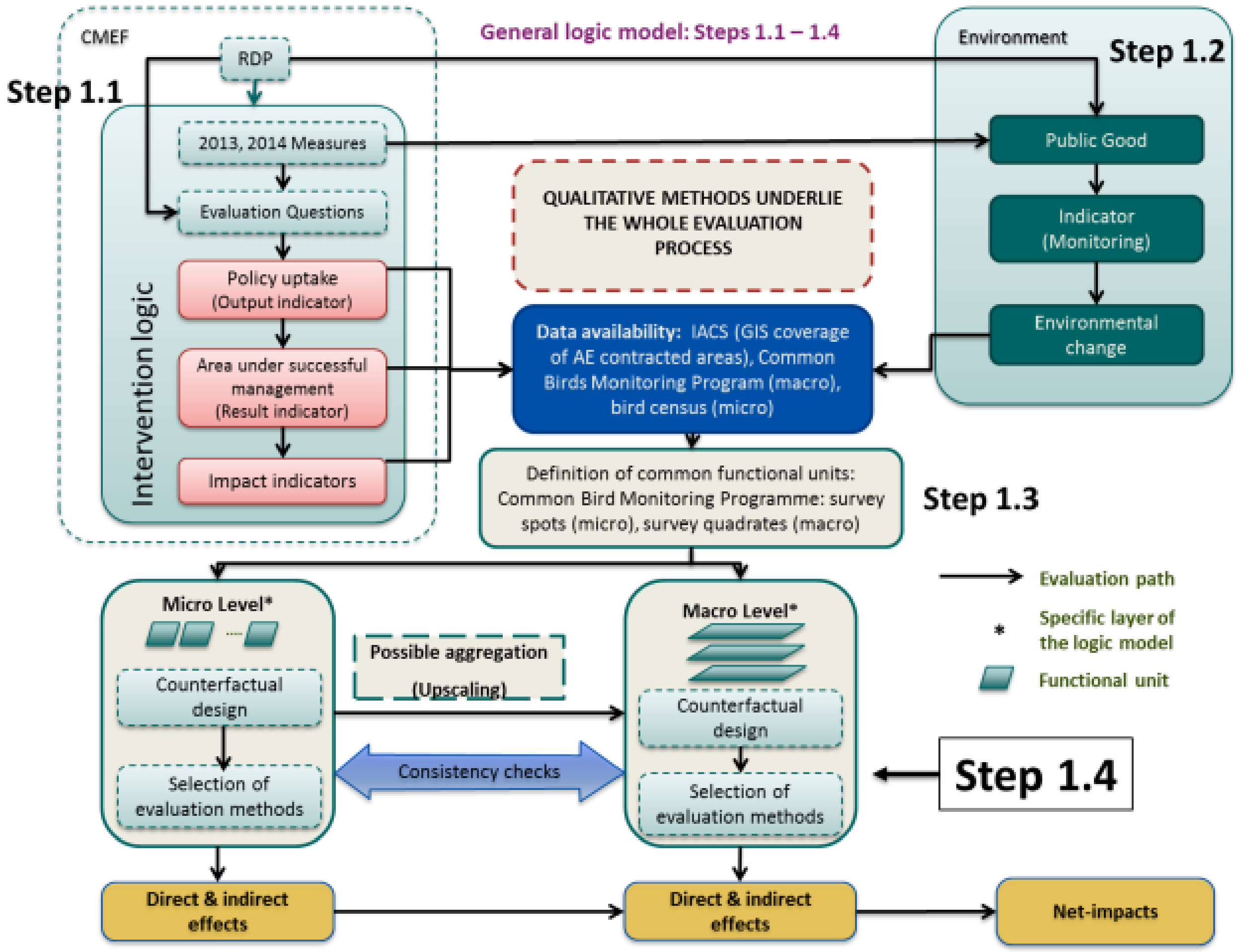

Figure 30: **Step 1.4** Counterfactual micro and macro-level evaluations and net impact assessment

Workflow and description of the counterfactual design

Step 2.1 & 2.2 – Input to the counterfactual logic model and defining comparison groups

Macro level: During the definition process of comparison groups the ratio of contracted areas under key policy measures was taken into account as the basis for grouping. As the evaluation is based on spatial statistics, functional units were classified based on the spatial coverage of the key policy measure (e.g. 214 agri-environmental measure). A total of 591 survey squares were assessed and, based on the percentile of coverage, three groups were formulated (non-participant, average participation, high participation). The three groups were assessed in parallel over a timescale of 1999-2014.

Micro level: During the definition process of the comparison groups of micro-level case study a slightly different approach was tested. 285 survey spots were classified based on the coverage of contracted parcels under the AE measure. Three classes were defined based on Jenks Natural Breaks method, and the class in the middle was ignored in further analyses, as the impact of the measure was expected to be hidden. Besides the assessment of the survey spots based on the uptake of the measure, further assessment was carried out concerning the ratio of natural areas within the survey spot. During the classification procedure the former HNVA designation methodology was used to identify natural areas based on CORINE 1:50.000 (Belényesi, 2007). Micro level comparison groups were:

1. Non-participant - natural
2. Non-participant – non-natural
3. Participant – natural
4. Non-participant – non-natural

The aim of involvement of an additional environmental attribute to comparison group design was to filter out the buffer effect of naturalness within the survey spots.

Step 2.3 – Selecting counterfactual-based evaluation options

For macro-level evaluations, the three groups were assessed in parallel from 1999 to 2014, while at micro level these groups were compared between 2009 and2014.

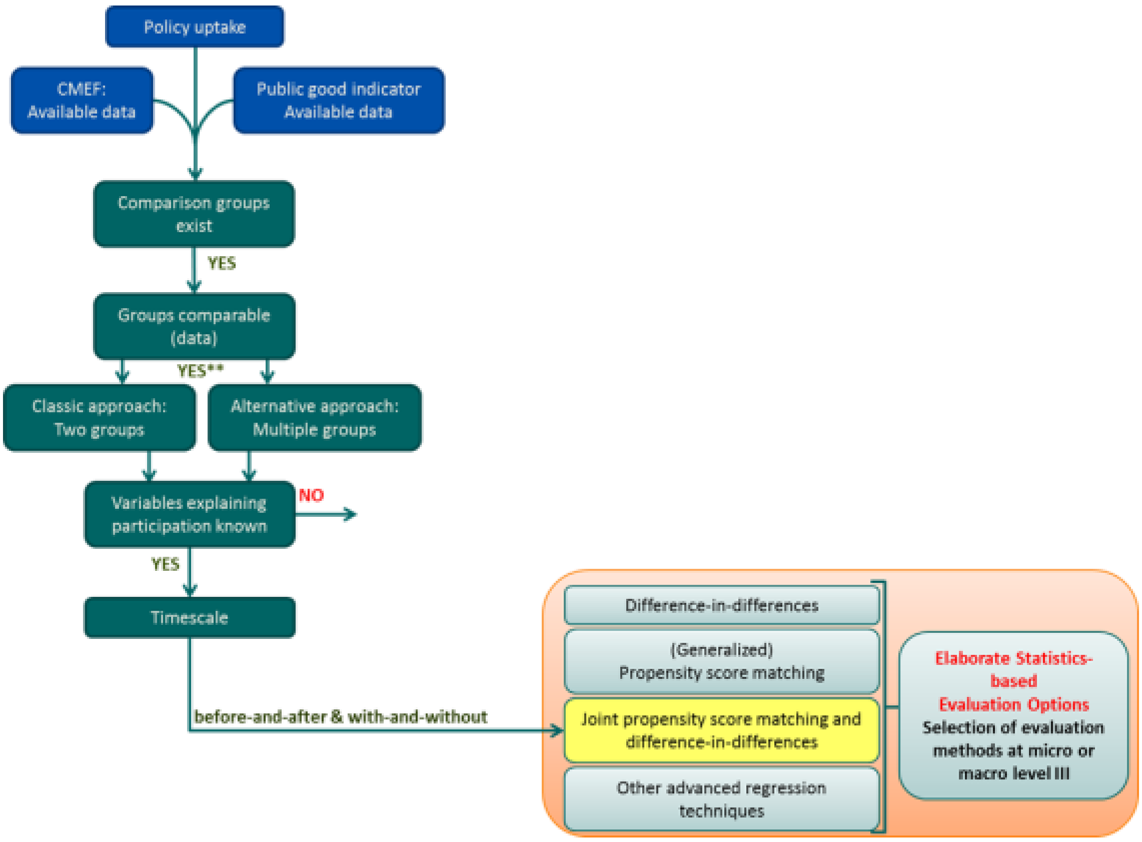

Figure 31: **Step 2.1, 2.2 and 2.3** Counterfactual design at micro and macro level

Steps of the micro level evaluation: including contribution of the approach to assessing net impacts

Step 3.1. Definition of functional unit and the consistency of the indicators.

Micro level: Data sets available in 285 functional units (survey spots) from 2009-2014.

Step 3.2. Assessment of data quality

Measure groups with sufficient sample sizes are included in the assessment providing a good data quality.

Step 3.3. Application of long run evaluation options w/o comparison groups

Micro level: Two-way ANOVA analysis was carried out to find potential differences during the survey period and the groups formulated. Results showed no significant differences in the number of species during the years, but highlighted significant differences among the number of species of the formulated groups. Duncan range test was also carried out which showed that the highest numbers of species can be observed in the non-participant-natural category, followed by participant-natural, participant-non-natural and non-participant-non-natural categories.

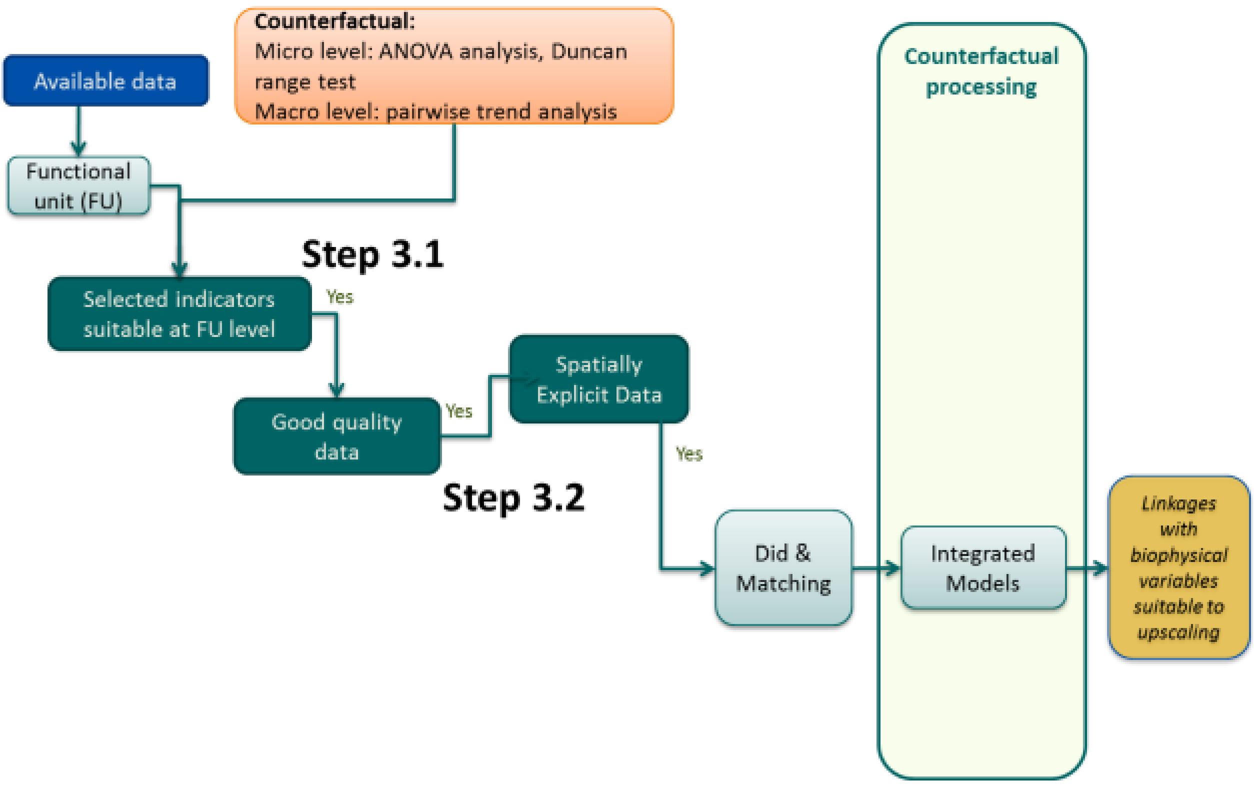

Figure 32: Macro and micro level logic model of the FBI and number of bird species indicators

Steps of the macro level methods: including contribution of the approach to assessing net impacts

Step 4.1 Definition of functional unit and the consistency of the indicators

Data sets available for functional units (survey squares) from 1999-2014.

Step 4.2 Assessment of data quality

Measure groups with sufficient sample sizes are included in the assessment providing a good data quality.

Step 4.3 Application of long run evaluation options w/o comparison groups

The values of the FBI were calculated based on the geometric mean and the standard error of the estimation of the geometric mean of the population index of the bird species in concern changes (Gregory et al. 2005). Comparisons were made in terms of the different comparison groups designed, and pairwise trend analyses was carried out in terms of participant, non-participant and average participant groups.

4.2 *Application of logic models for evaluation focussed on Climate Stability public good in Italy*

The climate issue is one of the biggest environmental challenges in the coming decades and climate stability is one of the main environmental objectives of the new legislative proposals for CAP reform. In particular EU Member States have set themselves the goal of reducing greenhouse gas (GHG) emissions by 20% of 1990 levels by 2020. The EU was also required by the Kyoto Protocol to decrease GHG-emissions to 8% below the 1990 levels in the period 2008-2012. Agricultural emissions have an important role, considering that methane and nitrous dioxide (almost the only GHGs coming from the farming sector) account for 10% of Europe's GHG emissions and carbon sequestration provided by Land Use Land Use Changes and Forestry (LULUCF) gives an important contribution to the net emission balance, especially in Italy. For this reason, climate stability is a very important public good to be evaluated in the performance of rural development programs.

Climate stability can be evaluated with the carbon footprint approach (CF) in terms of GHG balance (emission and sequestration), considering that this indicator accounts for all of the GHG emissions by the agricultural sector. The unit of measurement of CF is the equivalent tons of carbon dioxide. CF includes 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).

Description of case study region

Emilia Romagna is located in the northern part of Italy; it stretches from the Apennines to the Adriatic Sea and covers a big part of the River Po low-plain. The regional territory occupies about 22.500 km^2, and it is about 48% low-lying, 27% hilly and 25% mountainous. The population is about 4,430,000. The Po Valley is one of the most intensively farmed areas in Italy. Emilia-Romagna is a leader region for Italian agricultural production. There are approximately 84,000 regional farms. Over the past decades, the agricultural sector increased its competitiveness through deep structural reorganisation that led to highly specialized and innovative production. The productive sector in agriculture has on the one hand strictly territorial roots oriented in typical and high quality production, and on the other large-scale industrial production for trade.

Most relevant policy measures

The main Measures related to this public good are included in Axis 2 and concern GHG emission reduction from agriculture, carbon sequestration from LULUCF and the production of energy from renewable sources. More precisely, a specific goal of the RDP is the GHG emission reduction, mainly to be achieved through the maintenance or increase of the carbon stock in the soil and the reduction of input and energy demand. The main Measures are the following:

- 214 contains the main sub-measures for climate change, in particular sub-measure 214/A (Increase carbon storage in woody biomass); 214/B (Preservation and storage of carbon in soil); 214/E (GHG emissions reduction; nitrous oxide from chemical fertilisers);
- 216 (Support for non-productive investments)
- 221 (Increased renewable energy production)

More than 208,000 hectares (20% of the regional UAA) have been involved in actions for the environment. The agri-environment Measure 214 is mainly focused on organic production (59,000 ha) and integrated production (56,000 ha).

Table 1 State of implementation of the Emilia Romagna RDP

	RDP measure	UAA (ha)
211	Natural handicap payments to farmers in mountain areas	78.800
212	Payments to farmers in areas with handicaps, other than mountain areas	16.451
214	Agri-environment payments	157.535
216	Non-productive investments	150
221	First afforestation of agricultural land	5.065
226	Restoring forestry potential and introducing prevention actions	1.079
227	Non-productive investments	661

Application of the logic models

Approach 1 - Carbon footprint at process level

Steps of the general layer: CMEF, selection of environmental indicators, functional units, database concept and data sources

The Carbon Footprint approach has been implemented on different activities and sectors measuring the amount of carbon dioxide emissions that are directly and indirectly caused by an activity or accumulated over the life stages of a product at process level.

Step 1.1. following the CMEF intervention logic

The impact indicator of CMEF 2007-2013, aimed to measure the 'contribution to combating climate change', focused attention on the production of renewable energy with a common impact indicator defined in Guidance note J of the Handbook on CMEF as the 'Increase in production of renewable energy'. Among the baseline indicators on the assessment of 'quantitative and qualitative change in the production of renewable energy', there are also the GHG emissions from agriculture as ancillary information to the main indicator.

In the new CMES 2014-2020 the impact indicator linked to climate change is represented by 'GHG emissions from agriculture' as:

- Aggregated annual emissions of methane (CH_4) and nitrous oxide (N_2O) from agriculture;
- Aggregated annual emissions and removals of carbon dioxide (CO_2), and emissions of methane (CH_4) and nitrous oxide (N_2O) from agricultural soils (grassland and cropland).

The emissions are reported by Member States at national level respectively under the 'Agriculture' and 'Land Use, Land Use Change and Forestry (LULUCF)' sectors of the national greenhouse gas inventory of United Nations Framework Convention on Climate Change (UNFCCC). Emissions of CO_2 from the energy use of agricultural machinery, buildings and farm operations, which are included in the 'energy' inventory under UNFCCC, are not included in this indicator. This choice could lead to some underestimation of the effective contribution of RDP measures to the reduction of GHG emissions in the case of measures primarily focused on energy efficiency improvements, for example.

The analysis of the GHG emission at process level aims to contribute to measure the impact of an agricultural activity (e.g. productive process). The study is based on the analysis of productive process and related RDP measures by the Emilia Romagna Region during the period 2007-2013. The case study wants to explore the impact of agri-environmental measures on the reduction of GHG emission from agriculture.

Step 1.2. selecting additional environmental indicators

GHG emissions are well synthesised by the CO_2 equivalent indicator as it is widely used in reporting activities about climate change.

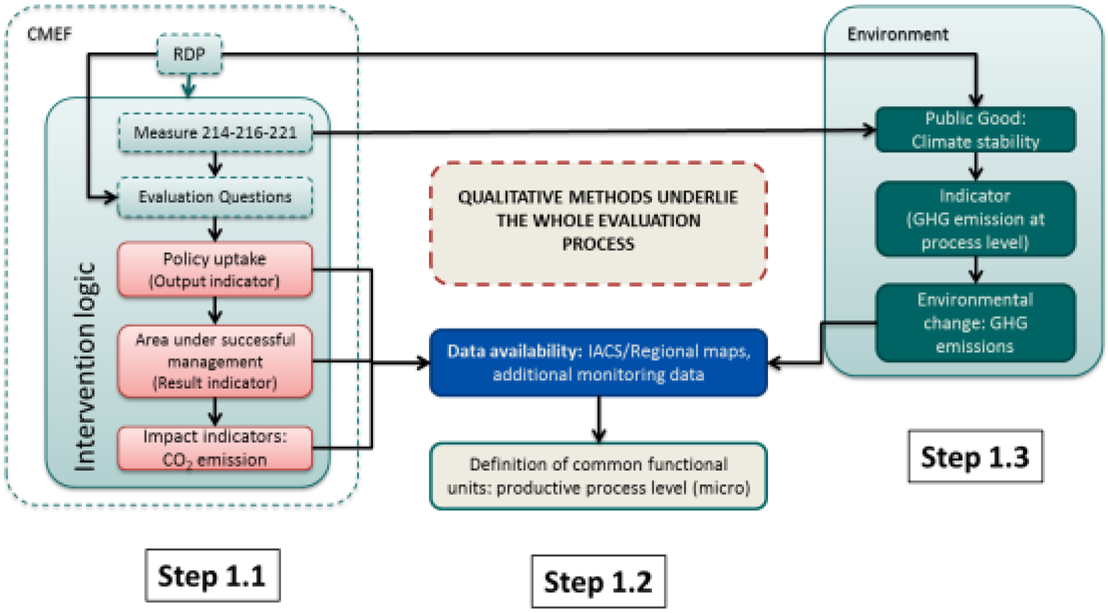

Figure 33: **Steps 1.1, 1.2, 1.3** of the general logic model

Step 1.3. selecting common functional units for micro and macro level evaluations

Data source for the assessment of carbon footprint at process level have been extracted from the IACS and LPIS, land maps and additional monitoring data (collected by the evaluator). Additional data were taken by the International Panel on Climate Change (IPCC), the Italian National Inventory Report (NIR) for GHG emissions and by interviews carried out on a selected number of farms. Milk yield data were provided by the National Agricultural Paying Agency.

The identification of the functional unit is important to set the boundaries of the production system (process level). In this case the CF analysis can be carried out starting from the raw product level (e.g. milk, grains, etc.) at farm boundaries or at final product on the market, considering in this case also the processing and transport phases. The functional unit proposed in this approach is the production process level (boundaries) and the emissions are proportional to the unit of production at crop and livestock level (Agriconsulting, 2013; Vitali et al., 2014).

Farm data were provided by interviews carried out on a selected sample of farms by the evaluator (Agriconsulting, 2013). The sample consists of more than 700 farms (factual and counterfactual) and 18 livestock farms. The data were directly measured and integrated with other data collected from different databases such as Nitrates Directive database (National Livestock Inventory), National GHG Inventory Report (NIR), IACS for crop distribution and data on the milk production from the National Bovine Registry, RDP 2007- 2013 register of Emilia Romagna and other official sources. Data referred to the years 2009-2011.

Step 1.4. Conceptual decision on counterfactual micro and/or macro level evaluations

Comparison group formation is particularly important when there is self-selection for programme participation. The use of counterfactuals and the construction of control groups are related to data from participants and non-participants. The control group design depends on the availability of data required at process level ('production unit') for participating and non-participating farms. Considering the data availability in the case study area through the use of different data sources (IACS, regional maps, additional monitoring data), a naïve estimate of counterfactuals (with-vs-without approach) has been carried out. The analysis of GHG emissions at process level by the comparison of programme participants and non-participants may suffer from sample selection biases and the lack of time-related factors leads to naïve estimators. 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.

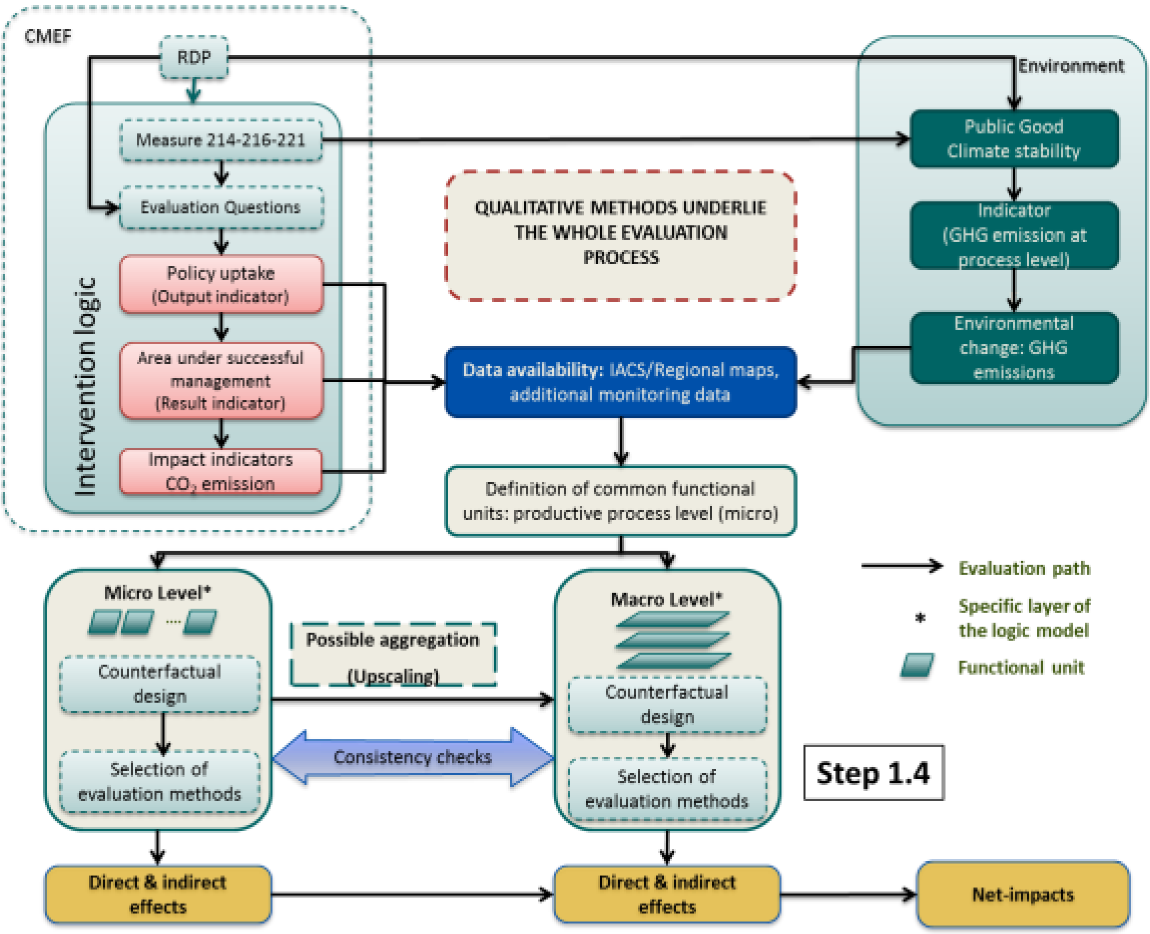

Figure 34: **Steps 1.4** Counterfactual micro and macro level evaluations and net impact assessment

The analysis at crop level was carried out with the aims: (i) to estimate the total CO_2eq emissions in six cropping systems (wheat, corn, alfalfa, pear, tomato and vineyard); (ii) to estimate differences in CO_2 emissions per hectare resulting from specific RDP measures (Organic Farming, Integrated Production and Advanced Integrated Production) compared to conventional production systems; and (iii) to aggregate the results of the analysis of the six cropping systems through the assessment of total CO_2 emissions for different RDP measures. The analysis for the livestock system was carried out with the following aims: (i) to estimate methane emissions (CH_4) from enteric fermentation and methane and nitrous oxide from manure management (CH_4 and N_2O) per livestock unit and kg of milk in selected and representative livestock farms participating in the organic farming measure of the RDP compared with conventional farms; (ii) to assess the net change of GHG emissions due to the participation at the RDP measure at regional level. The analysis evaluated the effect of the conversion of dairy cow and cow-calf farms from conventional to organic farming system in terms of greenhouse gas emissions.

Steps of the counterfactuals and comparison group design differentiate between counterfactual and comparison groups used at micro level and macro level

Counterfactual design at micro level

Step 2.1 Input to the counterfactual logic model

The RDP measures analysed in terms of GHG emissions are represented by Organic Farming, Integrated Production and Advanced Integrated Production. The first one was analysed for both cropping and livestock systems while the other two sub-measures was assessed only in the case of cropping systems. The CO_2 emission values of sites with agri-environmental measures are compared to the values of sites without any measures.

Information on farming practices and other characteristics was gathered by means of a questionnaire for each cropping system (soil tillage, irrigation methods, crop residue management, other farming operations, plant protection treatments, fertilization, type of machine used). The energy content was defined using values from the scientific literature allowing calculation of the emissions per hectare. Farm operations were classified on the basis of specific information (e.g. working depth, soil strength, number of operations, etc.) and evaluated in terms of GHG emissions. In the case study, the model has been tested to evaluate carbon emissions (CO_2) in different agricultural contexts (type of farming, geographic distribution).

Step 2.2 Defining comparison groups.

The classic two comparison groups are used to compare AEM participants with a control group represented by conventional farms. A comparison of factual (RDP participants) against counterfactual was based on two different functional units:

- the cropping system (wheat, corn, alfalfa, pear, tomato and vineyard);
- the livestock system (dairy cow, cow-calf and fattening beef).

The analysis was carried out on a sample of 700 parcels (factual vs counterfactual) with the collection of the main characteristics without considering emissions from buildings, activities of administrative management, farm position in relation to services and market place.

Information on the livestock farm demographic structure (cows, heifers and calves consistencies) was obtained from the National Bovine Registry. The 21 selected representative livestock farms are distributed among: six organic dairy farms, six conventional dairy farms, three herds of organic beef cattle, three herds of conventional beef cattle and three conventional fattening farms (without any factual farms due to technical limitation imposed by the organic farming disciplinary).

Step 2.3 Selecting counterfactual-based evaluation options

Within these data it is then searched if natural comparison groups arise, e.g. participants and non-participants. Due to the lack of farm structural data, there are no variables available that could explain the participation. For this reason a naive pairwise comparison is conducted with the stratified samples at process level.

A statistical analysis with the use of primary and secondary data collected was carried out by the evaluator. After this preliminary analysis, a method for the counterfactual design was made on the basis of the data availability. Due to limitation on data availability and lack of some variables in few cases the Naïve method 'with-without intervention' was selected.

The data are examined for variables that may statistically explain participation to the measure. The evaluator addressed specifically sample selection issues, where comparison groups may differ by population type due to different underlying qualities. Variables that explain participation to the evaluated measure or policy are case-specific and depend on the functional unit (e.g. regional vs. farm uptake of a measure). For regional analysis, this level of analysis may be difficult. The next step is to assess the timescale of data. The timescale can be based on cross-section data (with-and-without), i.e. data for the end of the evaluation period only, or data gathered both at the beginning and end of the evaluation period. Each step brings a new method for statistical analysis. Situations where there are no variables explaining participation or where only ex-post data is available lead to naïve approaches in evaluation.

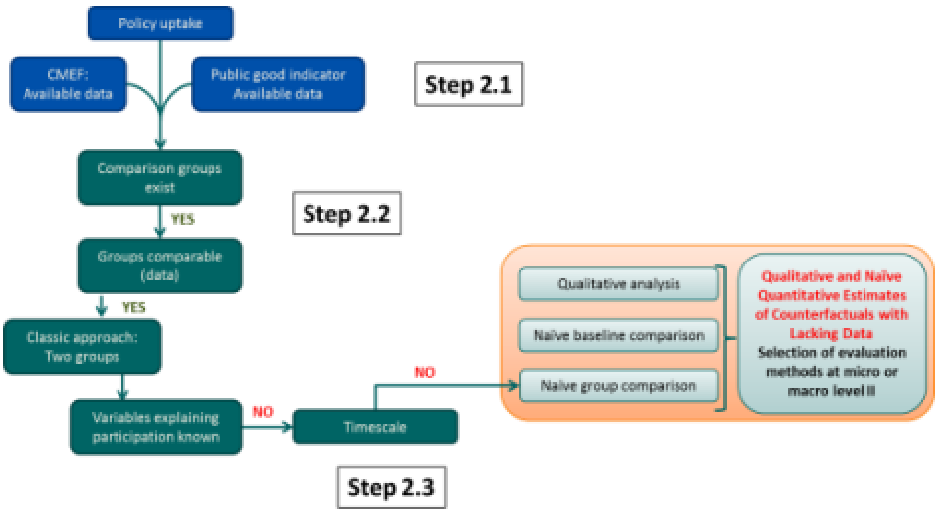

Figure 35: **Steps 2.1, 2.2 and 2.3** Counterfactual design at micro level

The comparison groups could differ in terms of observed and unobserved variables, and the simple comparison between beneficiaries and non-beneficiaries could result in distorted real net effects. Therefore it is very important to identify some meaningful micro-based policy parameters using available data on units in a given sample. For these reasons the formation of a good sample represents a substantial element to derive various parameters such as sample average treatment effect, sample average treatment on treated in order to be able to estimate aggregated impacts at the macro level by upscaling.

The process-level approach is based on the sample collecting in 2009-2011, from over 700 farms for representative cropping system (67% of regional crop in term of UAA distribution). The counterfactual approach was the naïve method with-without. The net effects were determined comparing the gross results of beneficiaries (factual) and non-beneficiaries (counterfactual). There are 2,828 combinations

(1,414 pairs factual / counterfactual) at the cadastral parcel level. Each beneficiary was combined with a neighbouring (counterfactual) farm falling on the same sheet of the cadastral map. For 331 parcels of a beneficiary farm, an equal number of parcels of non-beneficiary farms is added, distributed all over the region Figure 35.

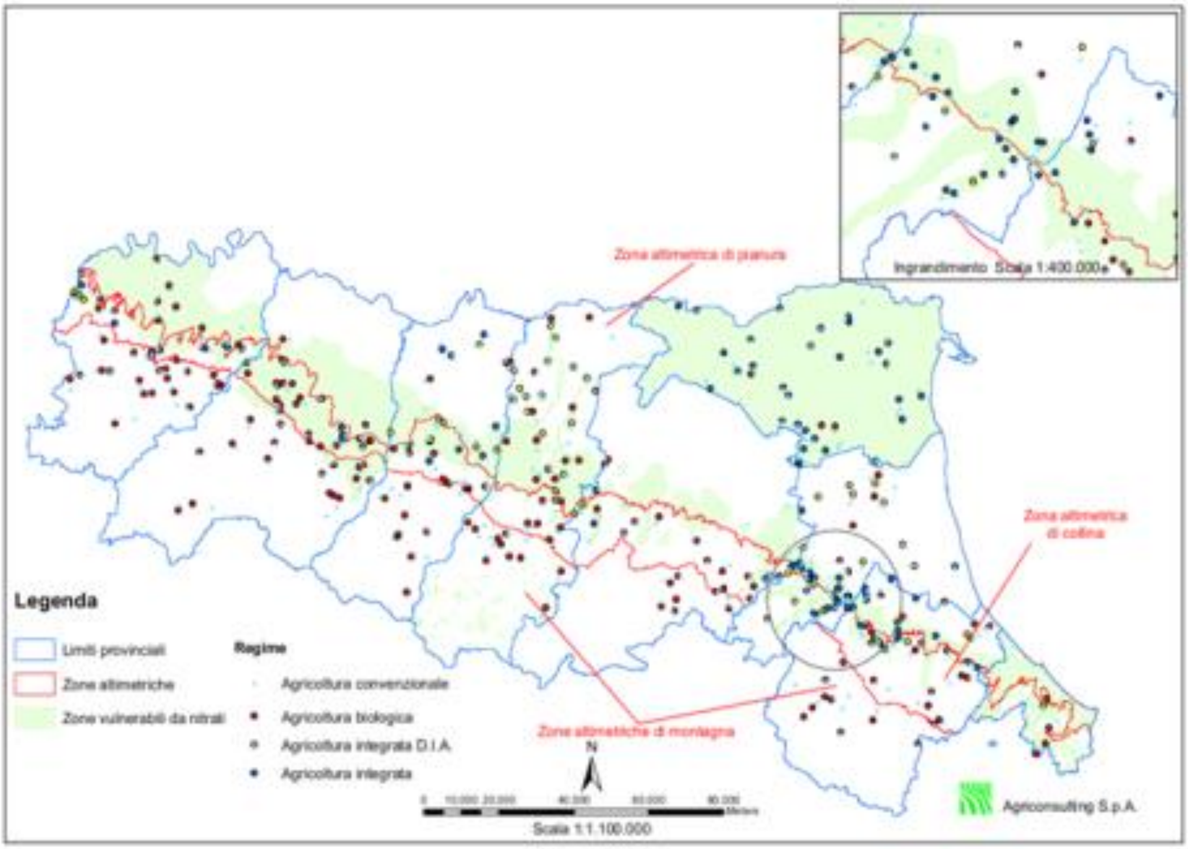

Figure 36: Regional distribution of sample areas to survey the process level functional unit

Steps of the micro-level methods: including contribution of the approach to assessing net impacts

The group of beneficiaries and non-beneficiaries was selected by comparing the gross results obtained by the beneficiaries of the programme (factual) with non-beneficiaries (counterfactual). For each beneficiary arable farm (factual) was combined with a neighbouring arable farm (counterfactual) falling on the same sheet of the cadastral map. The survey conducted between 2009 and 2011 includes 2,828 combinations of arable farms. They have been identified 1,414 pairs factual / counterfactual.

Step 3.1. Definition of functional unit and the consistency of the indicator

Carbon Footprint (CF) analysis needs to specify the 'functional unit' defined as a single productive process, single farm, or product, etc. - where matter and energy flows have to be accounted for. In this case study the functional unit for the CF calculation is defined as the agricultural productive process.

The first step concerns the configuration of an input-output framework to account for the amount of resources used to obtain the single agricultural output during a production cycle. It is a very important phase where all the various operations have to be carefully identified avoiding double counting in case of resource re-use (fodder, manure, etc.). In this step it is important to define the production system used, taking into consideration the differences between arable and livestock systems. Emissions for each crop (fertilizers, mechanization, pesticide, etc.) need to be recorded considering also the carbon soil sink. For animal production the analysis focuses on the CH_4 emissions from enteric fermentation and CH_4 and N_2O from manure management. The accounting is carried out according to the quantity of inputs used during the production cycle of one year (Table 2).

Table 2 Overview of the farm inventory data according to emission sources

Emission source	Main GHG considered	Others gases considered
Managed soils:		
Direct:		
Mineral fertilizer application	N_2O	NH_3
Manure application	N_2O	
Crop residues (including leguminous)	N_2O	
Pasture	N_2O	
Cultivation of organic soils	N_2O/CO_2	
Indirect:		
N deposition of NH_3/Nox	N_2O	NH_3
Leaching and runoff of nitrate	N_2O	
Livestock:		
Enteric fermentation	CH_4	
Manure management (housing and storage)	CH_4, N_2O	NH_3
Energy use:		
Direct:		
Fossil fuels	CO_2	
Electricity and grid energy (heating system)	CO_2	
Indirect (processing and transportation of used inputs):		
Fertilizers and amendment	CO_2	
Feedstuff	CO_2	
Machineries and buildings	CO_2	
Pesticides and seeds	CO_2	
Other inputs (plastics, livestock fees etc.)	CO_2	
Carbon storage:		
From agricultural soils (stock of organic carbon) and impacts of practices	Organic C / CO_2	
From on-farm trees	Organic C / CO_2	

The methodology adopted is defined in the guidelines provided by IPCC (Intergovernmental Panel on Climate Change) and ISPRA (Institute for the Protection and Environmental Research) to prepare the national inventory of emissions (ISPRA, 2011). Greenhouse gases emitted were estimated using the approach suggested in the guidelines provided by the IPCC and in accordance with Italian National Inventory Report (NIR). The CH_4 and N_2O emissions were expressed as kg CO_2 equivalent (CH_4, kg x 25; N_2O, kg x 310) to account for the Global Warming Potential (GWP). The GWP (mean ± standard deviation) were referred to Livestock Unit (LU equivalent to 600 kg live mass) or to kg of milk (the latter for the dairy sector only).

Step 3.2. Assessment of data quality

In the second phase the assessments of the quality of data have been carried out in order to check if the amount and characteristics of data are appropriate to implement one of the models available for the impact evaluation. The limited data quality often affects the applicability of the models for the environmental impacts assessment steering toward lack of consistent, robust or representative results.

For dairy cows the production of methane/head/year is lower than those reported in the national inventory of emissions (ISPRA, 2011). Although the methodology adopted is essentially the same, these differences are attributable to the various quantitative and qualitative production values used in the two estimates and the different composition of the sample analysed (national and regional). For beef cows and calves (6-24 months) the comparison is not possible because of the different methodology used (estimate of raw energy from the needs for maintenance and growth in the present study and estimation raw energy from the levels of ingestion in ISPRA report) and a different classification of the animal categories. The emission factors calculated for methane from manure are in line with those reported by ISPRA. This stems from the fact that the volatile solids calculated in the two estimates were derived from the same set of data (database nitrates).

Step 3.3. Application of environmental footprint approach

The Ecological Footprint can be defined as the area of biologically productive land and water that an individual, population or activity require to produce all the resources it consumes and to absorb the waste it generates, using prevailing technology and resource management practices. More specifically, the Carbon Footprint (CF) has been implemented on different activities and sectors measuring the amount of carbon dioxide emissions that are directly and indirectly caused by an activity or accumulated over the life stages of a product.

CF represents a simple tool for assessing GHG emission performances of agricultural processes (crop and livestock systems) and allows us to:

- compare carbon emissions in different agricultural processes (farm types) and among different farms;
- evaluate differences among agricultural production methods considering several RDP measures compared to conventional production methods;
- infer regional results (macro level) to evaluate RDP environmental impact.

CF is a good indicator that lends itself well to the calculation of a single production unit (crop processing), representative as micro level. The GHG emissions can be expressed in tons of CO_2 equivalent (t CO_2) through the use of conversion coefficient based on the Global Warming Potential (GWP) of each gas. The whole GHG net emissions can be measured by the aggregate indicator defined Carbon Footprint (CF), considering the full life cycle of a process or product (British Standard Institute, 2011; Pandey and Agrawal, 2014; Cucek et al., 2012).

CF can be considered as a subset of data derived from Life Cycle Assessment (LCA). In order to compare different systems, some standards exist to apply this at international level. In particular, this refers to the introduction of new regulations published in 2013, containing 14,067 specific principles, requirements and guidelines for the CF quantification and communication of a product, based on International Standards on LCA for quantification (ISO 14040, ISO 14044) and on environmental labels and declarations (ISO 14020, ISO 14024, ISO 14025).

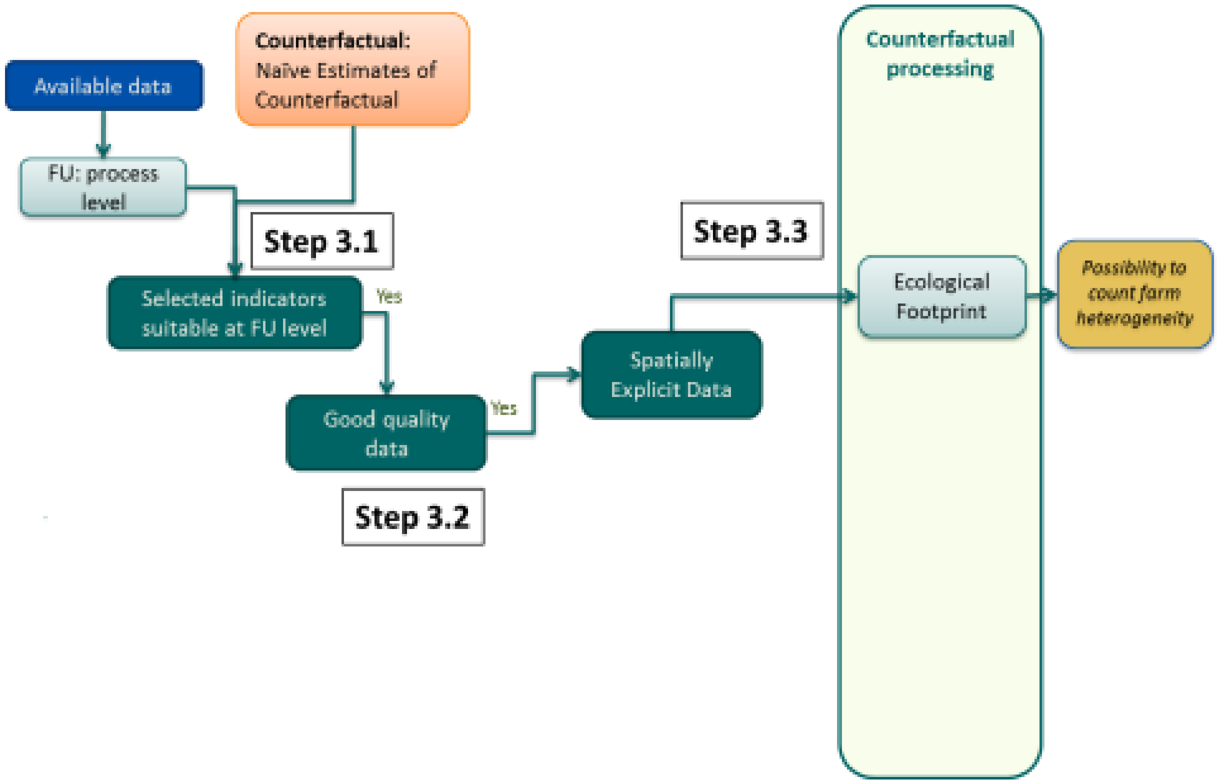

Figure 37: Layer of micro level logic model of GHG emission (Naïve Estimates of Counterfactual)

Steps of the macro-level methods: including contribution of the approach to assessing net impact

Step 4.1 Definition of FU and the consistency of the indicators

At macro level the functional unit of the CF approach is represented by the single process unit used at micro level. The evaluation of CF at process level is substantially an approach at micro level. In this case, to know the effects of the programme at the macro level (regional), it is possible to aggregate direct and indirect impacts computed at the micro level inferring from individual responses of farms.

Step 4.2 Creation of consistent (spatial) data

For upscaling of net effects, homogeneous zones have been identified on the basis of the main physical and environmental characteristics of the territory (Figure 38).

Step 4.3 Application of Naïve Estimates of counterfactual

Once calculated, the level of GHG emissions for each process (parcel level) this was weighted on the utilized agricultural regional area. The simple difference of these values, calculated in the factual and counterfactual, allowed then to estimate the overall effects at regional level (macro).

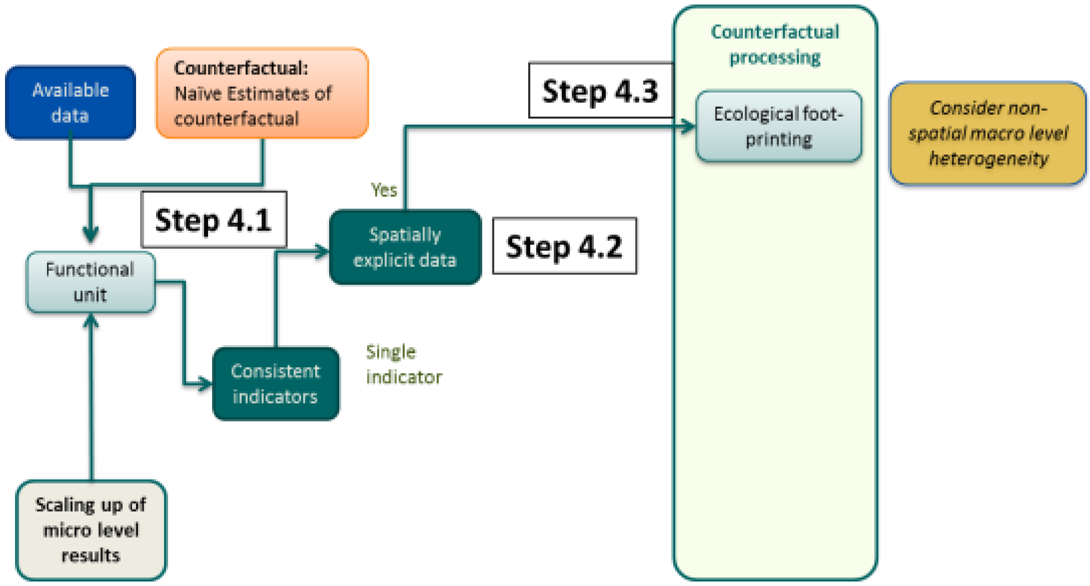

Figure 38: Layer of the macro-level logic model for GHG emission at process level (Naïve Estimates of Counterfactual)

Step of micro-macro linkages

The checks are designed to provide a comparative analysis of the data input used in relation to potential inconsistencies between the different statistical sources. The validation of the data should be carried out through direct monitoring. This allows us to overcome problems that might occur with data aggregated at the macro level for different types and agricultural systems. The micro-level data set is aggregated to the macro level to test the consistency of the analysis and validate the outcome of the results. Due to the limitations of the data, problems occur on data aggregated at the macro level for the different productive processes at regional level.

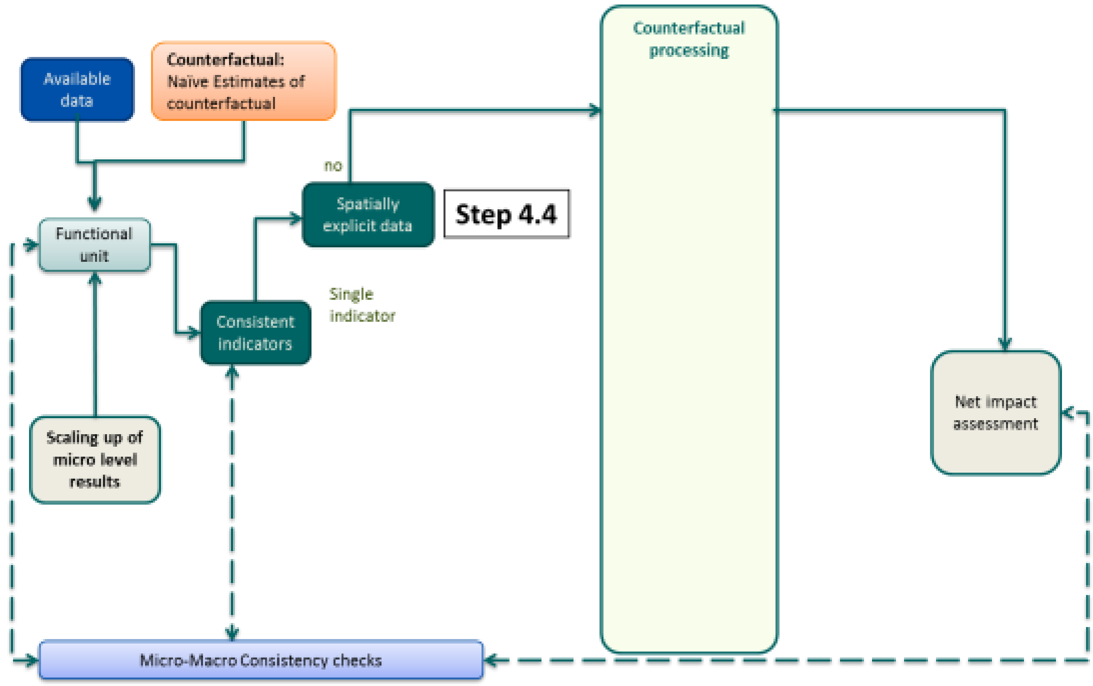

Figure 39: **Step 4.4** Micro-macro aggregation and validation

Approach 2 - Carbon Footprint at farm level

The second approach aims to calculate the direct and indirect GHG emissions considering the single farm as a functional unit. The calculation of direct and indirect GHG emission uses the Carbon Calculator (CC) created by a consortium coordinated by the EU Joint Research Centre, a user-friendly open-source tool designed to assess the life cycle GHG emissions from different types of farming systems across the whole EU (Tuomisto et al. 2013; Bochu and Metayer 2013).

The method is based on a modular approach, with the aim of estimating GHG emissions. More precisely, in the CC, the type of GHG emissions considered are: carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O) and hydrofluorocarbons (HFC).

The routes of the logic model do not differ from the one showed for the process-level approach. To follow the steps of the farm-level approach, see the figures used in the previous section.

Steps of the general layer: CMEF, selection of environmental indicators, functional units, database concept and data sources

Step 1.1. Following the CMEF intervention logic

For the introductory comments on CMEF see Section 4.1.1.

The preparation of the database was made mainly through the selection of several variables from the FADN dataset related to data input: (assessment registration; livestock processes; cropland processes; energy used) and additional data provided by direct surveys and questionnaires submitted on the farm sample. Considering the long time required in the calculation of emissions for each farm, the analysis was only carried out on a small number of representative farms on a regional basis.

Step 1.2. Selecting additional environmental indicators

GHG emissions are well synthesised by the CO_2 equivalent indicator as it is widely used in reporting activities about climate change.

Step 1.3. Selecting common functional units for micro and macro-level evaluations

The functional unit, identified by the system boundaries, is the farm in terms of 'ha of UAA' over a reporting period of one year considering all emissions from upstream of the farm (cradle) to the farm gate. Direct and indirect GHG emissions are considered, including emissions from the processing and distribution of inputs. The carbon storage in soils and in farm trees is also taken into account in the assessment.

The selection of a functional unit at micro level was done on the basis of the data availability and the possibility to calculate the CF at farm level. The functional unit of reference proposed in this approach is at the farm gate (farm system and its system boundaries) and the calculation of the emissions at farm level is proportional to the unit of UAA.

The control group design depends on availability of data required at farm level for participating and non-participating farms. Considering the potential data availability of FADN sample in the case study area, quantitative methods linked to quasi-experimental design could be applied. The use of different data sources (FADN, FSS, IACS and direct surveys) should also guarantee the analysis in the temporal dimension.

Step 1.4. Designing counterfactual micro and/or macro level evaluations

Comparison group formation is particularly important when there is self-selection to programme participation. Considering the potential data availability in the case study area through the use of different data sources (IACS, regional maps and additional monitoring data), only a naïve estimate of counterfactuals (with-vs-without approach) was carried out. Table 3 provides an overview of the farm inventory data according to emission sources, while several types of information can be derived from FADN database.

Table 3 Type of data at the farm level derivable from FADN database

Level	Type of data	
Farm ID	Farm code	Feed expenditure
	Latitude	Forage expenditure
	Longitude	Energy expenditure (aggregated)
	Farm type	Fuels expenditure
	UAA	Electricity expenditure
	Irrigated UAA	Irrigation water expenditure
	Livestock Unity (LU)	Fuels consumption (lit)
	Machinery (kWh)	Electricity consumption (kWh /year)
	Annual work units (AWU)	Renewable energy consumption (kWh /year)
	Fertilization expenditure	Renewable energy production (kWh /year)
	Pesticides expenditure	
Cropland	Crop	Energy consumption
	UAA	Type of fertilizer
	UAA irrigated	N (kg/ha)
	Production (ton)	P (kg/ha)
	Yield (ton/ha)	K (kg/ha)
	Rotation (Y/N)	Type of pesticide
	Intercropping (Y/N)	Treatments (n.)
	Total volume of water (m3/anno)	Tillage operation
	Type of irrigation	Fuels consumption (lit. /op.)
Livestock	Livestock type	
	Number of animals (average)	Type of forage
	Live weight (kg) (average)	Forage produced and consumed (ton dry matter/year)
	Number of animals (start year)	Forage purchased (ton dry matter/year)
	Live weight (kg) (start year)	Total cost forage purchased
	Number days on the farm	Type of feedstuff
	% grazing	Feedstuff produced and consumed (ton dry matter/year)
	Number of animals (end year)	Feedstuff purchased (ton dry matter/year)
	Live weight (kg) (end year)	Total cost feedstuff purchased
	Number animals sold	Manure management systems
	Live weight an. sold (kg)	% dry matter
	Number animals purchased	Manure production (ton)
	Live weight an. pur. (kg)	Manure used in the farm (ton)
	Milk production (kg/year)	Manure purchased (ton)
	Milk powder purchased for calves (kg)	Manure sold (ton)
	Cost animals purchased	

Steps of the counterfactuals and comparison group design, differentiate between counterfactual and comparison groups used at micro level and macro level

Counterfactual design at micro level

Step 2.1 Input to the counterfactual logic model

In case of poor data availability, a naïve estimate of counterfactuals (with-vs-without approach) has to be used. CF represents a tool for assessing GHG emission performances of agricultural processes (crop and livestock systems) and allows us to:

- compare carbon emissions in different agricultural processes (farm types);
- evaluate differences among agricultural production methods considering several RDP measures compared to conventional production methods;
- infer regional results (macro level) to evaluate RDP environmental impact

Step 2.2 Defining comparison groups.

At this stage, the counterfactual logic model starts with the data availability found in the general logic model. Two comparison groups are used.

Step 2.3 Selecting counterfactual-based evaluation options

Data from the previous stage are collected together and then searched if natural comparison groups arise, e.g. participants and non-participants. CO_2 emission values of sites with AEM are compared to the values of sites without measures.

At this stage the number of possible comparison groups has been decided. Then the data are queried for variables that may statistically explain participation in the measure. The evaluator addresses specifically sample selection issues, where comparison groups may differ by population type due to different underlying qualities.

The CF evaluation at the farm level is substantially an approach at the micro-level. In these cases, to know the effects of the programme at the macro level (regional), it is possible to aggregate direct and indirect impacts computed at the micro level using a quasi-experimental approach, starting from individual responses of farms. However the comparison groups could differ in terms of observed and unobserved variables, and the simple comparison between beneficiaries and non-beneficiaries could give outcomes distorting real net effects. Therefore it is very important to identify some meaningful micro-based policy parameters using available data on units in a given sample. For these reasons the formation of a good sample represents a substantial element to derive various parameters such as sample average treatment effect, sample average treatment on treated in order to be able to estimate aggregated impacts at the macro level by simple upscaling.

The farm-level approach is based on a small representative sample of the regional agriculture with the possibility of being annually updated. In the FADN field of observation there is a great diversity of farms, in terms of economic dimension and farming type, ensuring a highly heterogeneous field of observation. This situation is compatible with the case study because GHG emissions depend basically on the level of farming intensity (where the economic dimension may be a driver) and of course on the type of production carried out, while the spatial distribution is only a secondary aspect. Then, starting from a large and representative sample as the FADN the simple aggregation of direct and indirect impacts computed at the farm level could give a valid measure at macro level.

Steps of the micro level methods: including contribution of the approach to assessing net impacts

Step 3.1. Definition of functional unit and the consistency of the indicator

Carbon Footprint analysis needs to specify the 'functional unity' defined as a single productive process, single farm, or product, etc. - where matter and energy flows has to be accounted. In this case study the functional unit for the CF calculation is defined as the farm, using the hectares of UAA (Utilised Agricultural Area) as unit of measurement.

Step 3.2. Assessment of data quality

This assessment has been carried out in order to check whether the amount and characteristics of data are appropriate to implement one of the models available for the impact evaluation. The limited data quality often affects the applicability of the models for the environmental impact assessment steering toward lack of consistent, robust or representative results. The additional data collection depends on the type of farming. The data input priority concern the assessment registration, livestock, cropland and energy consumed (in the 'Other inputs' module). For a more detailed analysis, other modules can be considered (e.g. Farm buildings, Machinery, Secondary inputs or Natural infrastructures). In terms of data source, the main limits of FADN regard additional data requirements about farm practices (e.g. manure management, number of treatments, livestock management, management residues, specific tillage operations, number of days for grazing, liquid fossil fuel, other energy use, specific water use and consumption, renewable energy use, etc.). In these cases, integration with direct surveys is needed. The main limits are that the data collection is very time-intensive and the fact that CC needs to be tested on field case studies in a diversity of farming systems in order to determine its robustness and reliability. Moreover the current version of the Carbon Calculator does not include any database for comparison of its results.

In this case study, the CF is valued at farm level (micro) and potentially it can be inferred at the regional level (macro) in order to measure the impact between RDP beneficiary and non-beneficiary farms. The main advantage of this indicator is the immediacy with which it is possible to assess environmental impacts in relation to GHG emissions. However, the CF analysis at farm level is complicated by the large amount of data required for LCA. Data cannot always be available and accessible on the basis of existing statistical sources, and therefore they need to be complemented by direct monitoring. Furthermore it is not always possible to obtain results with a good level of approximation, in particular in terms of carbon removal from soil (LULUCF). On this basis, an important challenge was represented by the identification of representative farms in order to define a benchmark in relation to the most common production methods. Information of farming practices is needed for a robust assessment, particularly for inferring results at regional level.

FADN database does not provide the following required data:

- distribution of effective input consumed for specific operation. In these cases, it was determined the total value of water, fuel and electricity but not for individual cropping process or operation (pumping, cars, etc ..)
- type and quantity of each fuel consumed (petrol, gas, etc ..)
- quantity of renewable energy produced and/or consumed;
- presence of natural infrastructures (grassing, ecological areas, etc ..)

- energy consumption of buildings, materials and single machinery
- consumption of seeds (values derived from technical manuals)
- number of pesticide treatments (value derived from technical manuals)
- for the vineyards the required values are in hectolitres per hectare of wine
- limited information on meadows and pastures (e.g. discrimination between pastures > or <30 years)

Step 3.3. Application of Naïve Estimates of counterfactual

The Ecological Footprint (EF) can be defined as the area of biologically productive land and water an individual, population or activity requires to produce all the resources it consumes and to absorb the waste it generates, using prevailing technology and resource management practices. In particular, the Carbon Footprint has been implemented on different activities and sectors measuring the amount of carbon dioxide emissions that are directly and indirectly caused by an activity or accumulated over the life stages of a product.

Steps of the macro-level methods: including contribution of the approach to assessing net impacts

Step 4.1 Definition of Functional Unit (FU) and the consistency of the indicators

At macro level CF can be used through consistent application of up-scaling of available farm level data. This includes the testing of scaling methods based on the interpolation-sampling and aggregation of data.

Step 4.2 Creation of consistent (spatial) data

Due to the high aggregation level, small samples sizes and the lack of farm structural data, data quality and representativeness at regional level the quality of the upscaling is poor. The analysis provides only basic information on the data set and results.

Step 4.3 Application of Naïve Estimates of counterfactual

The analysis at macro level does not allow sound statistical analysis due to limited samples size and amount of variables.

The micro-level data set could be potentially aggregated to the macro level to test the consistency of the analysis and validate the outcome of the results. There are several consistency checks to be made between the different sources of data at the micro level and between the micro and macro-level data. The checks are designed to provide a comparative analysis of data input in relation to potential inconsistencies between the different statistical sources. Due to data availability limitation, problems occur on data aggregated at the macro level for different productive process at regional level.

Step 4.4 Step of micro-macro linkages

There are several consistency checks to be made between the different sources of data at the micro level and between the micro and macro-level data. The checks are designed to provide a comparative analysis of data input in relation to potential inconsistencies between the different statistical sources. The validation of the data should be carried out through direct monitoring data. Due to data limitations, it may be necessary to define some representative farms to derive reference data for consistency checks. This allows us to overcome problems that might occur on data aggregated at the macro level for different types and agricultural regions.

4.3 Application of logic models for evaluation focussed on Landscape Quality public good in Greece

The selected case study focuses on objective set in the RDP of Greece to protect landscape, and in particular to maintain the pruning and propagation practices in vineyards on the islands of Santorini and Thirasia. At the present time the CMES does not include specific indicators for assessing the effects of RD measures under focus area 4A on landscape. Therefore the case study aims to address the lack of suitable impact indicators, in the CMES and relevant evaluation documents, to estimate the impact of agri-environmental measures on landscape through the alternative indicators 'land use change' and 'visual amenity'. Moreover, the case study will also contribute to the consideration of diverse place-specific environmental characteristics in the impact assessment, since the examined measure is implemented in a very limited and defined area with unique landscape characteristics.

Description of case study

Santorini or Thira is an island located in the southern Aegean Sea, at the island group of Cyclades (Figure 40). It forms the southernmost member of the Cyclades islands, with an area of approximately 73 km^2 and a 2011 census population of 15,250 inhabitants. It is 128 nautical miles from Piraeus and 63 nautical miles from Crete (www.thira.gr/epixeirisiako-programma.html).

Santorini as well as the islands of Thirasia and Aspronisi are remnants of the volcanic island Strogili. Strogili was a volcanic cone and its central part was blown up along with the volcano's crater by the Minoan eruption which took place in 1613 BC and resulted in the creation of what we now call the caldera of Santorini (Economou et al., 2007). The view from the side of the volcano is rocky and steep in contrast to the smoothness of the soil in the rest of the island (**Error! Reference source not found.**). Santorini is characterised as fairly flat with Profitis Elias Mount to be its higher peak in 567m. Parts of the volcano are: Nea Kameni (1707-1711 AD), Palea Kameni (46-47 AD), the submarine volcano Columbo (active) (1650 AD), the Christiana islands (www.thira.gr/epixeirisiako-programma.html).

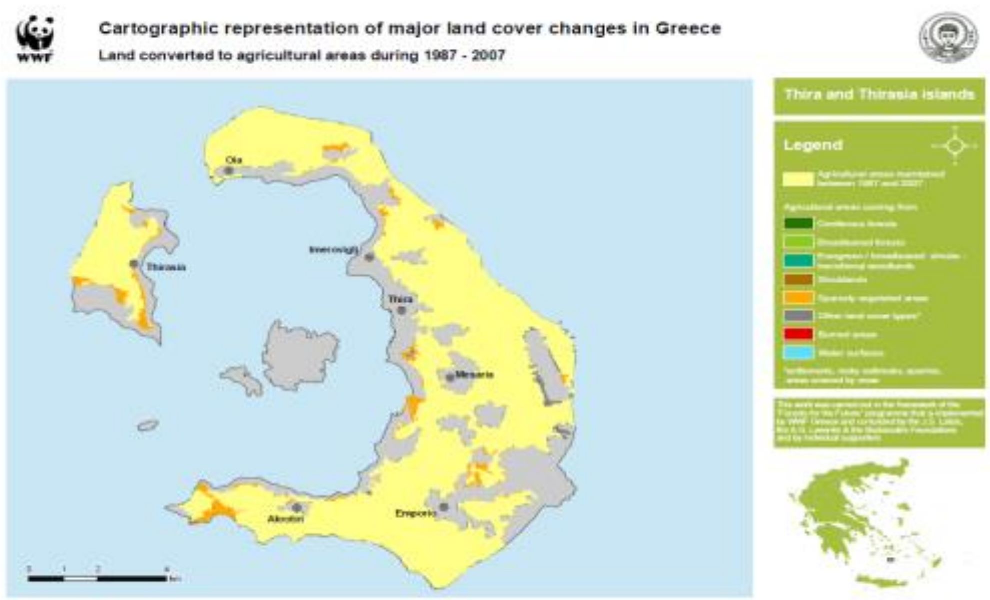

Figure 40: Map of Santorini island (Source: www.wwf.gr)

Generally, the economy of the island is based on the following sectors:

- Agriculture - fisheries
- Manufacturing
- Services
- Infrastructure - constructions

Agriculture was the main economic activity for the population of the island until 1970, when tourism emerged as the dominant economic sector. Farmers preferred to be engaged in tourist businesses and related activities that provide larger, easier and more stable income.

Agriculture-Farm structure

Agriculture sector in Santorini is characterised by extensive, i.e. low input farming, very specialised towards either viticulture or specific small scale tomato and Spanish vetchling (*Lathyrus clymenum*, Leguminosae) cultivation with minimum mechanisation (Rousou, 2006).

Apart from the vine cultivation, another significant crop in Santorini is a local variety of small size tomato that appeared in the island around 1875. Its variety is small fruited, bushy with thick peel, sweet and acidic taste at the same time, which was easily adapted to the volcanic soil, the strong winds and high temperatures of Santorini island (Rousou, 2006).

In the beginning of the 20th century and up to the earthquake of 1956, tomato was cultivated in 1,200 ha and over 10 tomato processing factories were operating for the production of canned tomato juice (Danezis, 2001). Currently the production volume has decreased by 95% and only one tomato processing factory exists, which belongs to the Union of Cooperatives of Santorini's products.

The third agricultural product of importance in Santorini is Spanish vetchling or 'Santorini's fava' as commonly known. It seems that 'Santorini's fava' have been cultivated continuously for 4,000 years on the island of Santorini, since the same strain of vetch was found in the excavations of ancient settlements in Acrotiri (Sarpaki and Jones, 1990).

Farm structure survey data in 1999/2000 suggest that there were 717 farms with a total of 1,591.84 ha of farmland. The average area per farm was 1.94 ha and the average land parcel size was 0.488 ha. Out of the total cultivated land 61.24% is covered with vineyards (641 farm holdings) and 34.42% with annual crops, mainly by 'Santorini's fava' and barley (438 farm holdings).

According to data from the president of General Board of Directors of the Union of Agricultural Cooperatives of Santorini) [8] in 2013, the total areas covered by vineyards and cereals, mainly barley, are 1,300 ha and 200 ha respectively. During the period 2010-2013, the 'Santorini's fava' cultivated area has doubled its size and is estimated at 250ha, while out of a total area of 150ha under vegetables cultivation, 65ha are tomato plants.

Vineyards

There is a long tradition of winemaking on the island of Santorini. Historically the beginning of viticulture and wine production is placed at the end of the 5th millennium BC before the great volcanic eruption.

[8] http://www.peliti.gr/index.php?option=com_content&view=article&id=232:sandorini&catid=11:fitiko&Itemid=16

There is also evidence for transportation and trade of Santorini's wines outside the island during the prehistoric period (www.thira.gr/epixeirisiako-programma.html).

In 1920 vines coverage reached 3,500 ha, accounted for 84% of land under cultivation (Kourakou-Dragona, 1994). However the area of vineyards was gradually diminishing over years, for instance in 1970 vineyards covered 2,250 ha and in 1997 cultivated area was even less, amounted to just 1,492 ha. This loss in vineyards could be attributed firstly to the massive earthquake in 1956 that forced many residents to emigrate and secondly to the rise of tourism (Drosou, 2005). Since 1997, it seems that viticulture tends to stabilize its cultivated area. The evolution of vineyards area on Santorini island is represented in the figure below (Figure 41).

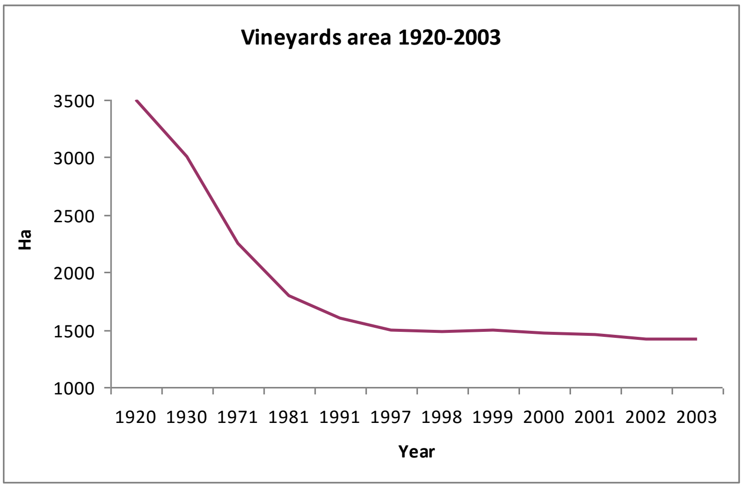

Figure 41: The evolution of vineyards area in Santorini 1920-2003

(Source: www.thira.gr/epixeirisiako-programma.html, elaborated by the research team)

The ecological environment of Santorini has remained the same for the last 3,700 years and the varieties of grapes -with the exception of few species- are indigenous to the broader Aegean region and adapted to the hot-dry climate and harsh winds of the island (www.thira.gr/epixeirisiako-programma.html).

Twenty-five indigenous grape varieties are grown in Santorini. The white grapes accounts for 80% of the total cultivated varieties, in particular Assyrtiko (70%), Athiri and Aidani (10%). The red variety Mandilaria covers 18% of the cultivated area, while the remaining 2% is comprised of many other varieties, e.g. Voudomato and Mavrotragano (Drosou, 2005).

Vines are self-propagated through layering in a disorderly manner in space and are planted at a distance of 2-2.5m from one another.

It should also be stated that Santorini's local grape varieties are totally resistant to phylloxera insect. Grape phylloxera (*Dactylosphaira vitifoliae*, Phylloxeridae) is a small, closely related aphid which feeds on the roots of vines producing galls that gradually strangle the circulation of the root and eventually the infected root shrivels and dies. It seems that volcanic geology of Santorini made its grape varieties immune to phylloxera, and thus Santorini remains one of the only places in Europe with its original un-grafted vines (Kourakou-Dragona, 1994).

The farmers are applying tradition methods of pruning – archaic approach so-called 'giristi or stefani' and 'kouloura'.

Currently the traditional vineyards of Santorini are being at risk. Firstly, producers and wine makers shifting towards a market and tourism oriented vine cultivation and wine making were the first that used modern techniques in farming, such as changing the pruning system into more intensive techniques, like the linear system. There are arguments suggesting that 'raising' the vines and supporting them on linear systems would enable farmers to lower the costs through mechanisation and also improve wine quality, since plant protection interventions would be far more effective. Hence Santorini wine and winemakers would be more competitive in the global wine market.

As also said before, the second pressure exerted on the vineyard landscape is that of urbanisation. Years of tourism development have left indelible traces on the landscape. Construction along the main roads and beaches and the expansion of urban construction around the main settlements, much of which is illegal, has created an urban continuum on a large part of the island. Dispersed construction outside this continuum has also contributed to a degradation of the landscape. For land owners, farmers included, the perceived opportunity cost of using the land for agriculture is extremely high. Hence, all previous attempts to control construction for tourism and leisure have been in vain.

Most Relevant Policy Measures

The case study examines Agri-environmental Measure (AE) 214, in particular sub-measure 4 Protection of agricultural landscape that was formed from agricultural activities (action 4.2 Maintenance of pruning and propagation practices in vineyards on the island of Santorini)

A local AE scheme was specifically designed for landscape protection on the two adjacent islands, Santorini (Thira) and Thirasia, that offered the maximum per hectare amount (900€/ha/year) permitted for permanent crops under Reg. EC/1257/99.

This AE action was first applied in 2005 under the Greek Operational Programme Rural Development-Restructuring of the Countryside (2000-2006) and also announced again in 2006 as part of the Greek RDP 2007-2013.

Eligible areas are those that covered with vineyards on the islands of Santorini and Thirasia and reached up to a total area of 1,500ha.

The specific objectives of this AE action are to:

1) maintain the traditional agricultural practices associated with the vineyards;
2) preserve the unique biodiversity and ecosystems of the volcanic island;
3) protect soil from erosion, and;
4) conserve the indigenous vineyard varieties

Farmers owning vines on the island receive 900€/ha/year, if they abide by the following commitments:

a) maintain the traditional pruning techniques of 'giristi or kouloura'
b) maintain the terraces within their vineyards
c) don't use mechanical or chemical weed control methods
d) create ecological compensation areas of 0.5m from each vineyard field margin edge

Thus the specific AE scheme compensate farmers for increased costs due to the maintenance of the specific pruning system and the terraces, as well as for revenues foregone due to decreased productivity, together with the protection of bushes and trees at the field margins.

The scheme had considerable success in terms of uptake and, during the first two years of implementation, almost half (47.27% out of 1,500ha) of the vineyards were under the scheme. According to the Greek mid-term evaluation document, there were a total of 655 beneficiaries and 709 ha of vineyards were supported.

Special measures in favour of the Small Aegean Islands concerning certain agricultural products (under Reg. EU 1405/2006)

Within the framework of a special aid to the small islands of the Aegean, taking into account their geographical and economic handicaps such as remoteness, insularity, small size, difficult topography and climate, support is granted for traditional activities, improvement of the quality and development of local products. This support was first introduced in 1993 under the Reg. EEC 2019/93 and it has been running onwards with considerable success ensuing that agricultural activities are maintained and quality products are supplied.

These special measures aim to limit the additional costs involved in transporting certain agricultural products to these regions (e.g. supplies with feeds and flours), and to encourage the development of local production of certain products. Currently the following agricultural products are subsidised:

- potatoes for human consumption and seed potatoes,
- tomatoes of Santorini island,
- plums of Skopelos island,
- beans (*Lathyrus sativus*, Leguminosae) as well as vetches (*Lathyrus clymenum*, Leguminosae),
- apiculture,
- mastic trees of Chios island,
- traditional olive groves,
- vines for the production of quality wines in traditional wine-growing zones,
- artichoke of Tinos island,
- citrus fruits,
- barley of Limnos island,
- milk for the production of traditional cheeses

Therefore, the scheme constitutes an example of the collateral protection of landscape as a ‘positive externality’ when aiming at the maintenance of agricultural activity.

In our case and as far as the vineyards are concerned, support is granted for the continued cultivation of the traditional vines of Santorini that produce quality wines, i.e. Protected Destination of Origin (PDO) and Protected Geographical Indication (PGI) wines. The aid is amounted to 700€/ha/year while the total estimated area covered by vines reaches to 5,200ha in the Aegean region. The aid proved effective since the tendency of land abandonment was avoided and furthermore cultivated area has been stabilised. In the case of Santorini, an average of over 80% of the vineyard received the support during the period 2002-2006 (Ministry of Rural Development and Food, 2008).

In general, one could say that the overall available amount of aid (1,600€ per ha), if the two schemes were to be adopted, could compensate for the increased costs of cultivation in the traditional way, but it is rather doubtful whether this amount could be enough to compensate farming households for the opportunity cost of agricultural land use. Therefore the implementation of these schemes also has clear socio-economic benefits for the island community focusing on the maintenance of rural society.

Application of the logic models (Description/explanation of the evaluation approaches along the steps of the logic model)

Approach 1: Land cover change

Steps of the general layer: CMES, selection of environmental indicators, units of analysis (selection of appropriate scales), database concept and data sources)

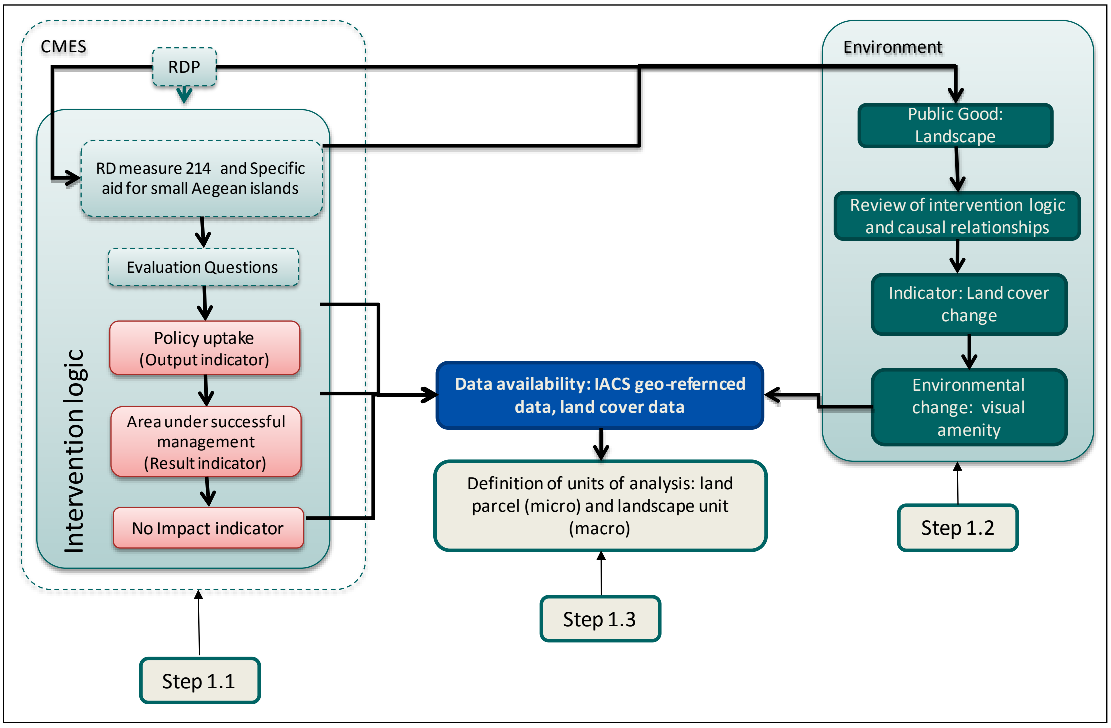

Figure 42: **Steps 1.1, 1.2 and 1.3** of the general logic model of the land cover change indicator

Step 1.1. following the intervention logic

The examined measures that the case study aims to explore are:

- the AE action for the maintenance of pruning and propagation practices in vineyards on the island of Santorini, and
- the special measures in favour of the Small Aegean Islands concerning certain agricultural products, and in the specific case of Santorini the maintenance of traditional vines that produce quality wines

The output indicator is the policy uptake (output indicator No34: Number of Farm Holdings and Holdings of Other Land Managers Receiving Support, output indicator No35: Total Area under Agri-Environment Support, output indicator No36: Physical Area under Agri-Environment Support, output indicator No37: Total Number of Contracts), the result indicator is the area under successful land management (result indicator of all measures under axis 2), while no impact indicator is mentioned in the CMES. Furthermore, according to the working paper of European Evaluation Network for Rural Development (2010), a list of additional landscape indicators was suggested by some Member-States, however none of these were suitable in our case (Table 4).

Table 4 Overview of the various approaches concerning the landscape by the Member States

Programme	Additional indicator	Measurement
Hessen	Landscape	Preservation and improvement of cultural landscape
Niedersachsen and Bremen	Landscape	Preservation and improvement of landscape coherence
Niedersachsen and Bremen	Landscape	Enhancement of landscape's cultural identity
Rheinland-Pfalz	Cultural landscape	As side effect with positive impacts of agri-environmental measures
Emilia Romagna	Maintenance and development of landscapes	Evaluation of: - coherence - differentiation - cultural identity
Scotland	Safeguarding the sensitive aspects of landscape character	(no measurement provided)
National	Preservation of attractive landscape	Attractive landscape will multiply by thousand ha

(Source: European Evaluation Network for Rural Development, Working Paper on Approaches for assessing the impacts of the Rural Development Programmes in the context of multiple intervening factors (March 2010))

Step 1.2. Selecting additional environmental indicators

On the right side of the first general layer, regarding the environment, the examined public good is the landscape, and the alternative landscape indicator 'land cover change' is selected.

This indicator is based on indicators developed in the IRENA operation (Indicator Reporting on the Integration of Environmental Concerns into Agriculture Policy), which is a joint exercise between several Commission Directorates-General (DG Agriculture and Rural Development, DG Environment, DG Eurostat and DG Joint Research Centre) and the European Environment Agency (EEA). The main purpose of IRENA operation was to develop agri-environmental indicators for monitoring the integration of environmental concerns into agriculture policy in the European Union (EU-15). Thus IRENA produced 35 agri-environmental indicators supported by data sets at NUTS 2/3 level (where data is available) and classified according the DPSIR model (Driving force - Pressure - State - Impact - Response). Among these proposed indicators, the pressure agri-environmental indicator No24 'land cover change' identifies land cover changes to and from forest/semi-natural and agricultural land.

In the Greek case study, this landscape indicator "land cover change" was chosen in order to compare changes in land cover, and in particular changes away from the traditional pruning practices to more intensive farming systems, at two different points in time using a spatial analysis. In our case, there is no environmental change but environmental effect, since the AE measure aims at maintaining the pruning and propagation practices in vineyards of Santorini. Thus the environmental change that is to be measured is how the landscape is perceived and experienced by people and the effect of changes on their visual amenity.

Step 1.3. Defining units of analysis for micro and macro level evaluations

The land cover change indicator is suitable for measuring changes at farm level, in particular it may detect changes in each land parcel. Therefore land parcel is considered the unit of analysis for the micro level analysis. The specific unit constitutes a level in three different scales i.e. a spatial, an administrative and a programme scale.

The aggregation of all land parcels forms the landscape of Santorini's island, thus regarding the macro-level analysis, the unit of analysis considered is the landscape unit, e.g. the whole landscape of the island. The macro unit of analysis pertains to all three scales mentioned above, spatial, administrative and programmatic.

Data sources

A. **Land cover data** of Santorini were drawn from satellite images. A GIS expert was assigned for mapping the land cover in Santorini, downloading satellite images from Google Earth (GE). GE releases free images taken at different periods of the year in high spatial resolution that may provide some potential for regional land use and cover mapping (Hu et al., 2013).

In our case study, freely accessible GE satellite images were used as direct data resources to detect changes in land cover of Santorini's area. For the island of Santorini, GE has three different images over time. The oldest image captured in 2003 (June 8), while the most recent in 2012 (July 8). One more image from 2010 is available, but it was not selected for our case study, since we wanted to have a bigger timeframe. Therefore the dates of capture, i.e. 2003 and 2012, fit well with the temporal dimensions of our examined measure. In particular, the AE action is implemented since 2005, thus this time frame of 9 years (from 2003 to 2012) enables us to study and estimate the land cover changes on the island of Santorini before and after its implementation. Due to time constraints, the testing of indicator focused only on the north part of Santorini, which includes the community of Oia.

B. **The IACS geo-referenced data** was freely provided by the Greek Payment Authority of CAP Aid Schemes (OPEKEPE: Payment and Control Agency for Guidance and Guarantee Community Aid). The available data set includes geographically referenced information on the field parcels participating in the AE action and special measures in favour of the Small Aegean Islands (including ha under the measures) in Santorini, only for the year 2011. Although data for participants and non-participants for the period 2007-2013 had been requested, this was not provided.

In particular, for the northern part of the island, data includes geographically referenced information on 102 land parcels that participate in the AE action covering 81.43ha vineyards area, and 23 land parcels that also participate in the special measures for the maintenance of traditional vines with a total of 15.64 ha of vineyards area. Data shows that all aforementioned field parcels also participate in the AE action for the maintenance of the pruning system.

Methodology

Images were first photo-interpreted and then manually digitized creating a geodatabase. Image interpretation is the process of examining an aerial photo or digital remotely-sensed image and manually identifying the features in that image. In general, manual interpretation typically involves delineation of polygon boundaries -areas with similar properties- and the subsequent classification of those polygons

(Morgan et al., 2010). A variety of key characteristics are used to delineate and classify polygons, including tone or colour, shape, size, pattern, texture, shadows, site, and context (Carleer and Wolff, 2006).

In our case study, for the techniques of photo-interpretation and manual digitisation, the GIS expert set the virtual GE eye[9] at 500-600m. This certain height above the ground was considered sufficient for detecting the different land cover categories. However specific landscape features such as terraces and boundary walls are not visible. Our focus was on the vineyards, paying particular attention to their pruning and planting system. As mentioned above, the traditional vineyards on Santorini are self-propagated through layering in a disorderly manner in space and are planted at a distance of 2-2.5m from one another. Thus, vineyards having traditional pruning system were easily distinguishable from vineyards that are supported on linear systems, using stakes or wires, because of their different spatial characteristics (e.g. size, shape, texture and pattern). Having also in mind the other crop types cultivated in Santorini along with its urban area, image interpretation produced datasets of polygons, which subsequently classified into six land cover types/categories. These datasets were the primary mapping products, the historical land cover map (2003) and the recent one (2012).

It should be stated that the historical land cover map was produced independently of the recent map, so as to avoid bias in previous interpretation and classification. This means that the classified polygons of the historical map were not used as reference polygons for the interpretation and classification of the recent map. This can be considered as a weakness of the methodological approach, to be corrected later on.

Furthermore, the historical GE image was unclear and fuzz in relation to the 2012 GE image. Therefore, the GIS expert couldn't clearly perceive the boundaries of the area around the caldera of Santorini (rocky area with cliffs) and delineate in accuracy the polygons. This limitation resulted in an overestimation of the bare land in the 2003 land cover map compared to the 2012 land cover map. As a consequence, it is observed that the total digitised land area in 2003 to be slightly bigger than in 2012 (a difference of 31.33 ha of land area). Given that the bare land is irrelevant to the purpose of our study, we assume that both land cover maps have the same ha of land area.

Preliminary classification consisted of:

1) traditional vineyards: vines grown in disorderly, almost at random locations and pruned with the traditional techniques of ''giristi or kouloura'
2) linear vineyards: vines trained horizontally along wires, which subdivided into two classes in relation to their planting density:
    a. in high density
    b. in lower density
3) other cultivated land: including all other crop types
4) bare land: mainly rocky areas without vegetative coverage
5) built up area: the settled area.

After this first classification of Santorini's area, a ground truth survey was conducted in order to ascertain the accuracy of the remote sensed data. Reif et al. (2012) argue that it is mandatory to have sufficient ground truth data in order to test the accuracy of the image analysis output. Therefore ambiguous and unclear satellite features that arose during image interpretation were selected as sampling points. Direct

[9] The GE Eye altitude represents the elevation of our viewpoint.

observations allowed us to test if what had been sensed was related to real features on the ground. Moreover, new GPS track data was collected during the field trip and imported into the GE. After correcting the interpretation mistakes, the preliminary classification of the land cover maps was adjusted.

At last, the following land cover classes were determined:

1) traditional vineyards
2) linear vineyards
3) area under annual crops
4) other cultivated area
5) bare land
6) built up area.

In particular, the previous two classes of linear vineyards, in high and lower density, were merged into one, since it was difficult to identify their planting density. Moreover, area under annual crops was used as an additional class, since the dates of GE images have captured in June and July, and by that time annual crops, such as barley and *Lathyrus clymenum*, have already been harvested. Therefore, in the satellite images, the class area under annual crops represents agricultural fields without crops at that moment. On the contrary, the class 'other cultivated land' includes permanent crops and spring/summer plants, e.g. trees and vegetables respectively.

The last step involved the integration of IACS geo-referenced data into the produced land cover maps. Data sets, IACS georeferenced data and land cover data, are both spatially explicit and at the same scale and resolution. Thus, in the historical and recent land cover map of Northern Santorini, the participating and non-participating land parcels were distinguished. Changes of each group between the two points over time, in 2003 and 2012, will be compared.

Step 1.4. Conceptual decisions on counterfactual micro and/or macro level evaluations

The AE action aimed at maintaining the traditional agricultural practices associated with the vineyards, preserve biodiversity, combat soil erosion and conserve indigenous vineyard varieties.

On the other hand the special aid for continued cultivation of traditional vineyards, intends to counteract the problems created to farming by insularity i.e. lack of accessibility, remoteness, adverse weather conditions etc., by, among others, supporting traditional farming activities.

However the pressures exerted on agriculture and Santorini's landscape, are of different origins but can be summarised to urbanisation, tourism developments and intensive farming.

The counterfactual scenario in the case of Santorini is built around these two types of pressures, endogenous (intensive farming) and exogenous (urban and/or touristic land uses).

While the special measure for the small Aegean islands is thought to play a role in reducing the pressure from urban/touristic land uses but not against intensive practices, in the absence of the AE and the incentives provided for maintenance, farmers would be expected to adopt the modern practices, since they would be cheaper and they could also increase their competitiveness. Nevertheless, one should not forget the fact that the incentive for maintenance of the traditional cultivation methods could also act as an incentive for the maintenance of agricultural activity.

Hence the hypothesis for the counterfactual is twofold. In the absence of the special measures for the small Aegean islands land use changes would be far more easy while if the AEM had not been implemented farmers would succumb to external pressures and even if they did not they would have an incentive to apply modern intensive techniques since the price for their product have been increasing due to increased demand.

Consequently, land cover changes will be visualised through the produced land cover maps and a comparison is going to be conducted among participants in AE action, participants in both measures, AE action and the special measures in favour of the Small Aegean islands, and non-participants (Table 5).

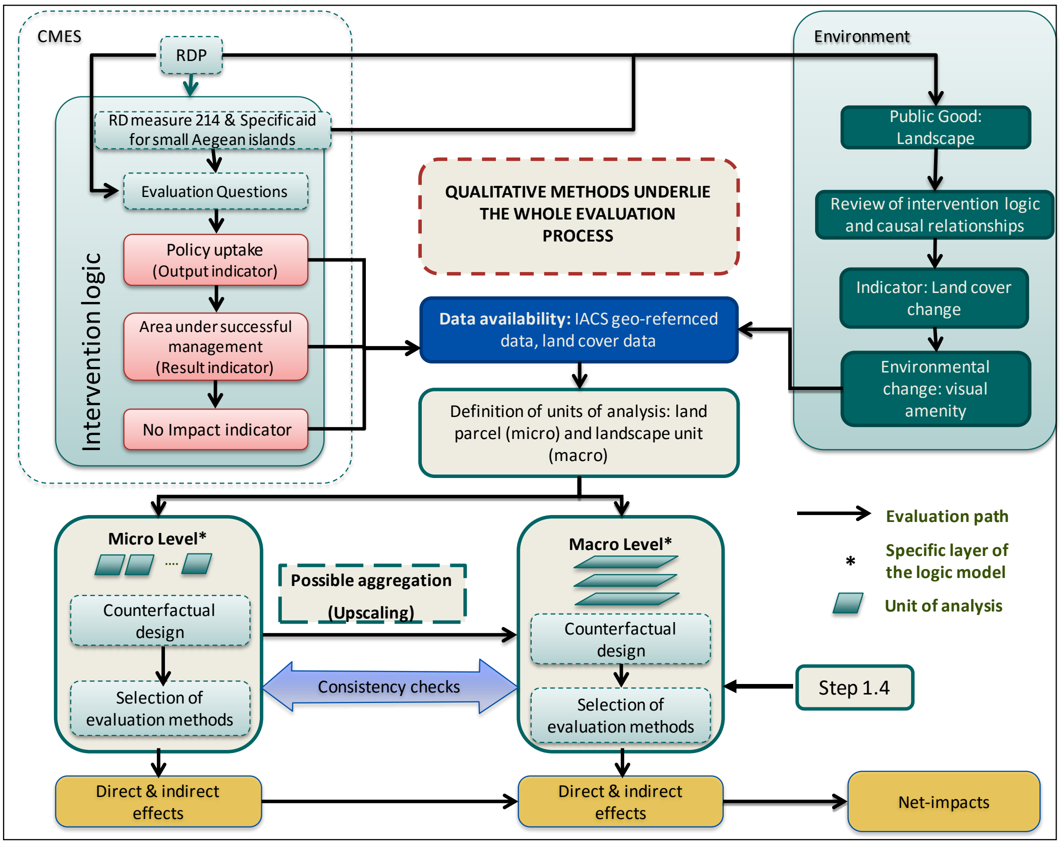

Figure 43: **Step 1.4** Counterfactual micro and macro level evaluations and net impact assessment of the land cover change indicator

Table 5 Comparison groups for the counterfactual

1st comparison group	2nd comparison group	3rd comparison group
Participants of AE action for the maintenance of pruning and practices in vineyards on the island of Santorini	Participants of AE action for the maintenance of pruning and practices in vineyards on the island of Santorini, and the special measures for the maintenance of traditional vines	Non-participants in either measure

Since the AE action applied in 2005 and land cover data are available for the years 2003 and 2012, the groups are comparable before and after measure implementation. Therefore the observed changes in land cover between the comparison groups will allow us to estimate the effectiveness of the policy intervention on vineyards landscape.

Workflow and description of the counterfactual design (Steps 2.1, 2.2 and 2.3)

Counterfactual design at micro level

Step 2.1 Input to the counterfactual logic model

In order to test the land cover change indicator, IACS georeferenced data and land cover data are required for participants and non-participants (see Figure 44).

Step 2.2 Defining Comparison groups

Data sets, IACS georeferenced data and land cover data, are both spatially explicit data and at the same scale and resolution, thus there is no need to include scaling mechanisms. Therefore, available data allow us to construct comparable comparison groups. There are three comparison groups, land parcels under the AE measure for the maintenance of traditional pruning practices, land parcels under the aforementioned measure and the special measure for the conservation of vineyards and land parcels under no measure. Particularly the latter group will determine our counterfactual scenario, the state of the environment without policy interventions.

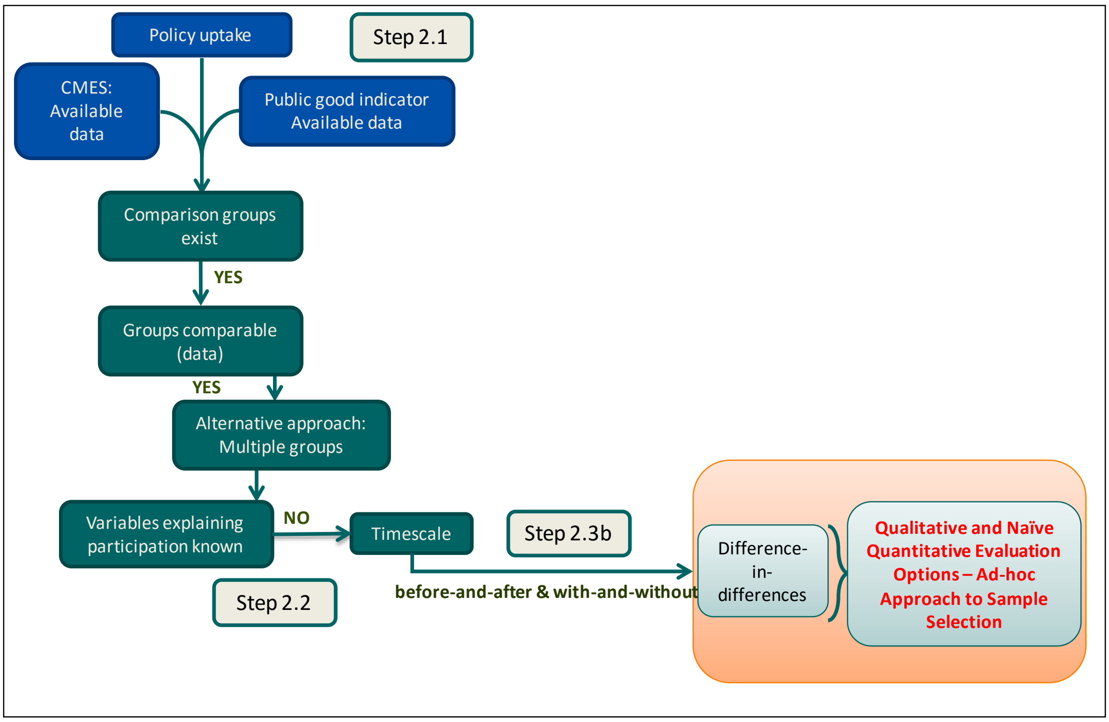

Figure 44: **Steps 2.1, 2.2 and 2.3** Counterfactual design at micro level of the land cover change indicator

Step 2.3 Selecting Counterfactual-based evaluation options

Although people who own vines on the island of Santorini are eligible for participating in both policy measures, data on comparison groups does not include variables that statistically may explain participation. Moreover, since our data sets cover the time period before and after the application of the AE action, this leads us to use the Difference in Differences in order to assess the effect of policy. A naive DiD approach compares the before and after situation of participants and non-participants in the programme. It requires data availability between two periods observed. In our case, the land cover changes between three different groups will be compared using a spatial analysis approach. Thus we will observe and measure changes or maintenance of land use and pruning/propagation practices within the timeframe 2003-2012 through land cover in:

- land parcels participating in the AE maintenance of pruning system,
- land parcels participating in the maintenance of pruning system alongside the maintenance of traditional vine cultivation, and
- land parcels participating in neither scheme

Furthermore, except for the vineyards, it is also interesting to observe the changes (losses, expansions) in area covered with the other land cover classes over the examined period on the island of Santorini (see Figure 44).

Counterfactual design at macro level

Step 2.1 Input to the counterfactual logic model

Land parcel level data will be aggregated, and analysed at landscape level. Aggregation would be the only upscaling approach to be used to link between the micro and macro level. Its consistency is attempted through the use of spatially explicit information (Figure 45).

Step 2.2 Defining Comparison groups

IACS georeferenced data and land cover data are spatially explicit. There are three comparison groups, land parcels under the AE measure for the maintenance of traditional pruning practices, land parcels under the AE measure and special measure for the conservation of vineyards and land parcels under no measure (see Figure 45).

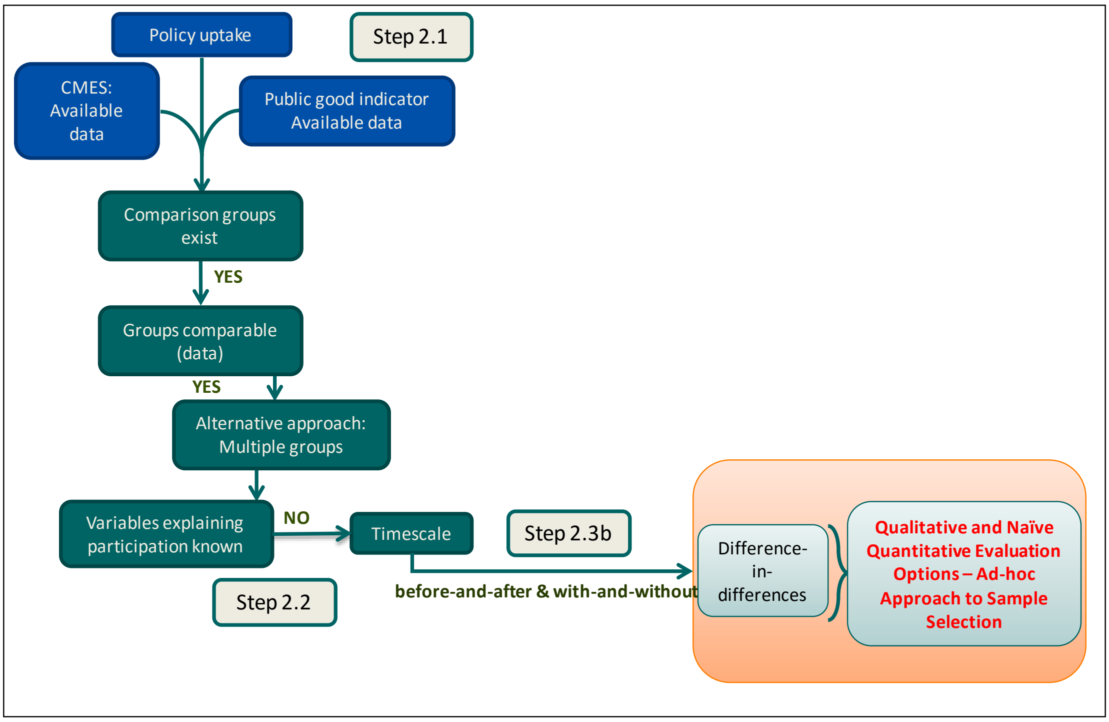

Figure 45 Steps 2.1, 2.2 and 2.3 - Counterfactual design at macro level of the land cover change indicator

Step 2.3 Selecting Counterfactual-based evaluation options

Although land users who own vines on the island of Santorini are eligible for participating in both policy measures, data on comparison groups does not include variables that explain participation. Moreover, since data sets cover the time period before and after the application of the AE action, this leads us to use a naive Difference in Differences counterfactual approach (see Figure 45).

Steps of the micro level evaluation: including contribution of the approach to assessing net impacts (Steps 3.1, 3.2 and 3.3)

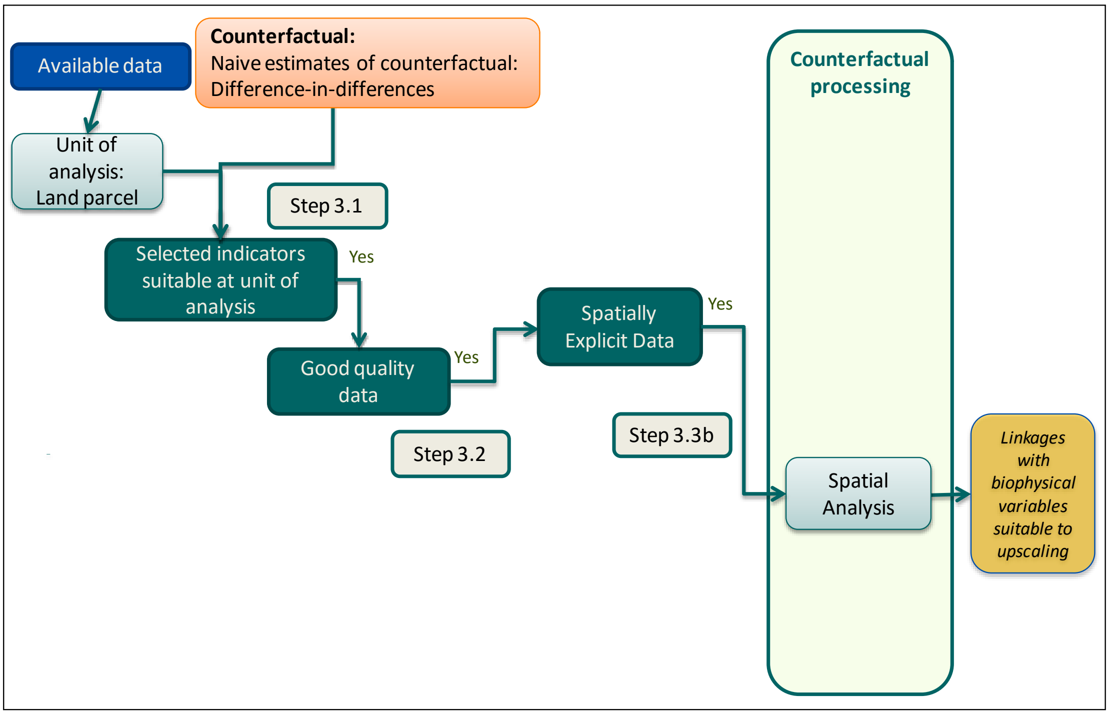

Figure 46: Layer of the micro logic model of the land cover change indicator (Naive estimates of counterfactual)

Step 3.1. Definition of unit of analysis and the consistency of indicators

IACS georeferenced data and land cover data are available at the land parcel level, while the selected indicator land cover change is also suitable at this micro level (see Figure 46).

Step 3.2. Assessment of data quality

IACS georeferenced data and land cover data are spatially explicit and of good quality. However information is not linked to the farm level, thus analysis doesn't allow for an advanced statistics-based impact assessment of policy measures.

Step 3.3b Application of naive estimates of counterfactual

Naïve estimates of counterfactual will be used, since data prior and after measure implementation are available, without using particularly complex modelling approaches. A change detection analysis will be used in order to identify changes in traditional vineyards that have taken place from 2003 to 2012 through polygon by polygon comparison within each comparison group (Figure 46).

We would attempt to estimate the indirect effects, such as deadweight effects, by checking change/maintenance observed in the land parcels non-participating in maintenance of pruning practices that would have occurred even in the absence of the applied measure, but information on the farm level was not available.

Steps of the macro level methods: including contribution of the approach to assessing net impacts (Steps 4.1, 4.2 and 4.3)

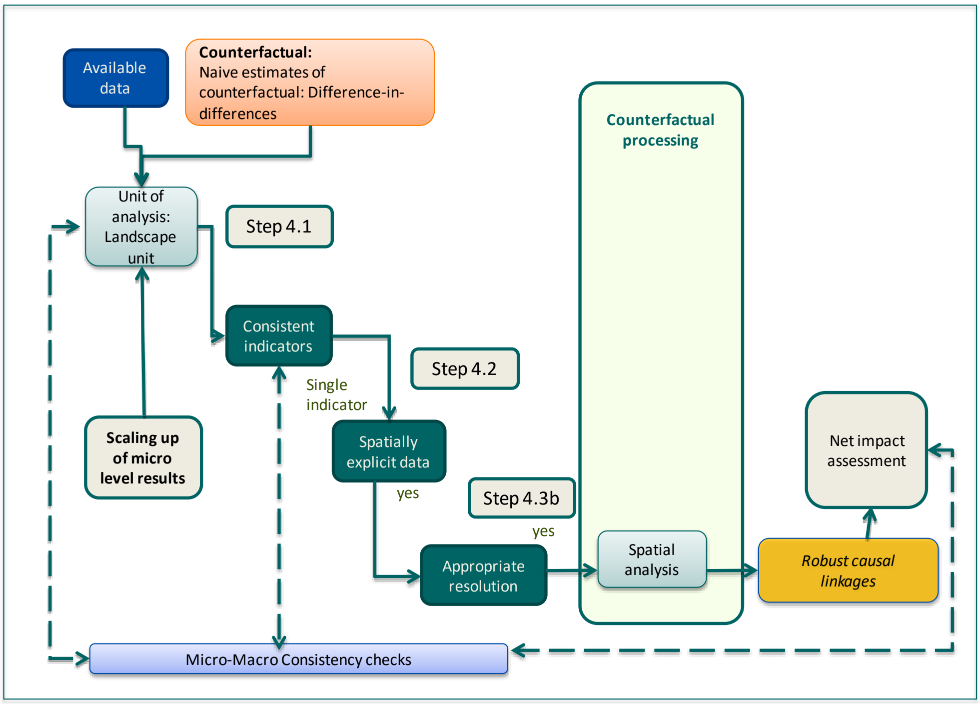

Figure 47: Layer of the macro logic model of the land cover change indicator (Naive estimates of counterfactual)

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

Land parcel level data will be aggregated and analysed at landscape level. Aggregation would be the only upscaling approach to be used to link between the micro and macro level. Its consistency is attempted through the use of spatially explicit information (see Figure 47).

Step 4.2 Creation of consistent (spatial) data

Land parcel level data will be aggregated and analysed at landscape level. We use a single consistent indicator that is spatially explicit. Data sets, IACS georeferenced data and land cover data, are both spatially explicit data and at the same scale and resolution, thus there is no need to include scaling mechanisms.

Step 4.3b Application of naive estimates of counterfactual

Since comparison groups are comparable and there is timeseries, but variables explaining participation are unknown, land cover change as a single consistent indicator will be used for the difference in differences assessment. Thus changes in traditional vineyards that have taken place from 2003 to 2012 through polygon by polygon comparison will be explored within each comparison group, trying to identify robust causal linkages between measures implemented and landscape changes. This spatially explicit approach has a great potential for improving the causality linkages between RDP measure and impact.

Steps of micro-macro linkages

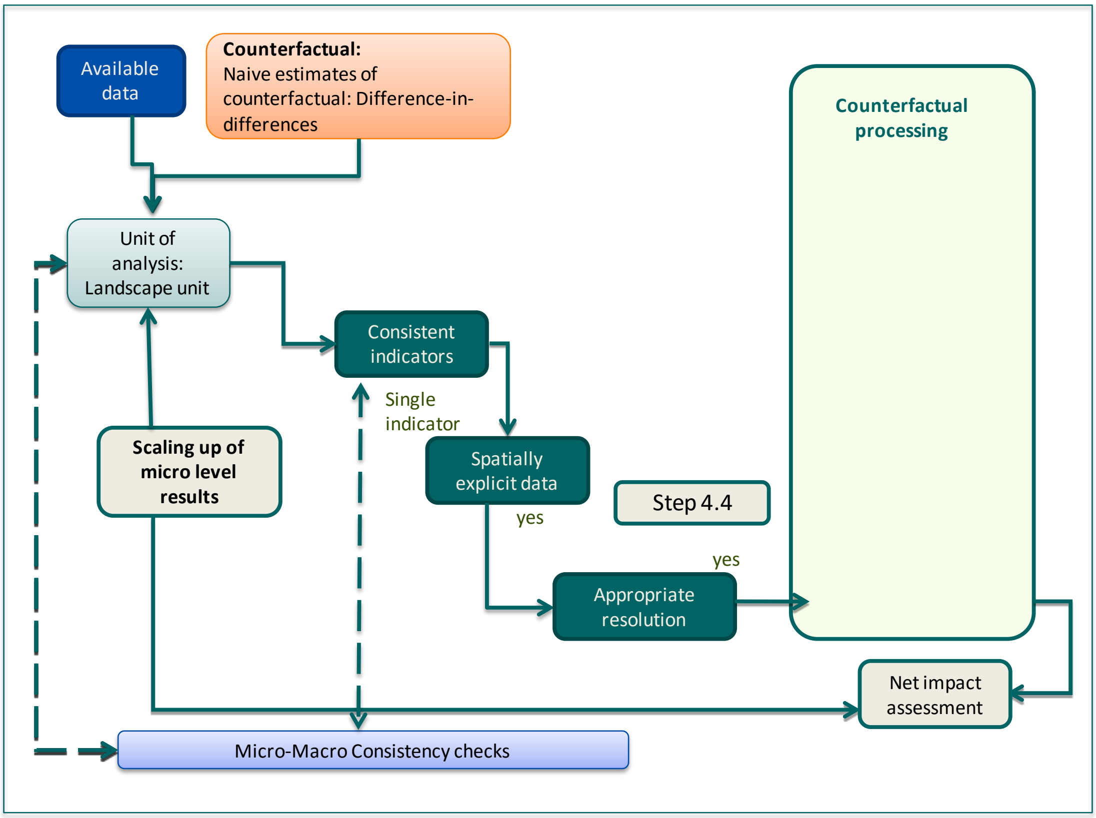

Figure 48: **Step 4.4** Micro-macro aggregation and validation of the land cover change indicator

Step 4.4 Micro-macro aggregation and validation

Land parcel level results form the basis of the macro level assessment. Results and micro level data from land parcel level will be up-scaled and aggregated to the landscape level.

We would attempt to estimate the indirect effects, such as deadweight effects, by checking change/maintenance observed in the land parcels non-participating in the action for the maintenance of pruning practices that have occurred without intervention, but lack of data at farm level inhibited such a comparison.

Approach 2: Visual amenity

Steps of the general layer: CMES, selection of environmental indicators, units of analysis (selection of appropriate scales, database concept and data sources)

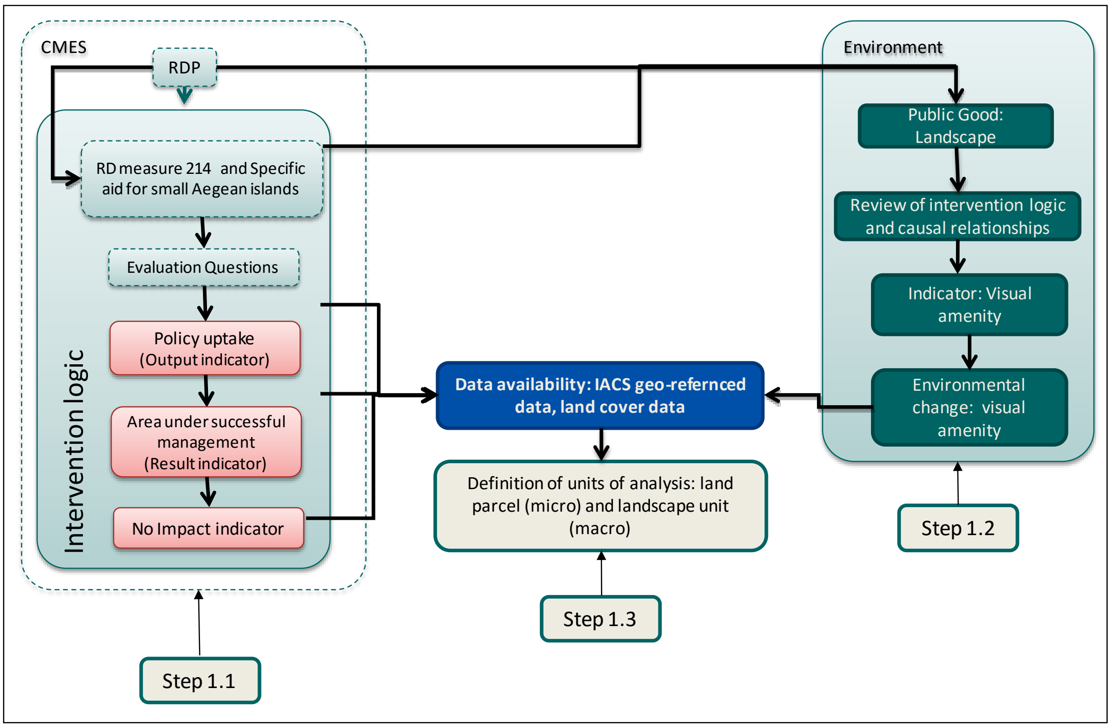

Figure 49: **Steps 1.1, 1.2 and 1.3** of the general logic model of the visual amenity indicator

Step 1.1. following the intervention logic

As mentioned above, the case study aims to estimate the effect of the followed measures on landscape:

- the AE action for the maintenance of pruning and propagation practices in vineyards on the island of Santorini, and
- the special measures in favour of the Small Aegean Islands concerning certain agricultural products, and in the specific case of Santorini the maintenance of traditional vines that produce quality wines.

However, currently there is no impact indicator included in the CMES to assess the impact on landscape.

Step 1.2. selecting additional environmental indicators

In our case study we included the amenity values arising out of Santorini's landscape with its traditional vineyards. According to the European Landscape Convention, landscape is an area, as perceived by people, whose character is the result of the action and interaction of natural and/or human factors (Council of Europe, 2000). Furthermore, as Tahvanainen et al. (2002) mention, there is a growing awareness that agriculture produces, usually as external effects of agricultural production, also environmental goods, like scenic beauty. Thus, the appearance of the landscape affects people's perception of their environment and the amenity that they enjoy.

Landscape visual quality can be defined as "*the relative aesthetic excellence of a landscape*" (Daniel, 2001) and examined in terms of observer appreciation (Lothian, 1999). Landscape structure and the presence of landscape attributes are important for the visual quality of the landscape and in shaping human landscape preferences (Tveit et al., 2006). Moreover, de la Fuente de Val et al. (2006) suggest that land cover structure plays a significant role in the visual quality of the landscape and landscape preference is related to the spatial distribution of landscape elements.

In many scientific studies, the scenic beauty of agricultural landscapes has been evaluated. The most common method for assessing scenic beauty of landscape is photographic representation (e.g. Arthur, 1977; Clay and Daniel, 2000; Wherrett, 2000). However, visualization methods, such as computer line graphics and computer-aided image-capture technology (ICT), have been also used to produce visual representations for evaluations of scenic beauty (e.g. Johnson et al., 1994; Tahvanainen et al., 1996; Tyrväinen and Tahvanainen, 1999).

In general, in order to assess the aesthetic preference or scenic quality of a landscape, human viewers are usually asked to make an ordinal ranking of several photographs or to assign a specific value (rating) to each of several photographs separately. The Likert-scale method developed by Likert (1932) asks individuals to express the extent to which they agree or disagree with a statement. The Analytic Hierarchy Process (AHP) introduced by Saaty (1977) derives a ratio scale of preferences (or priorities) using a set of pairwise comparisons. These comparisons are made in order to assess which of the two landscapes is more beautiful (scenic preference) and how much more beautiful. The disadvantage of this method is that it is not possible to evaluate large data sets. One solution for dealing with this limitation could be the approach developed by Alho et al. (2001).

The use of multiple regression analysis enables to relate statistically the landscape features to the scenic preferences of observers, establishing a mathematical relationship between the physical characteristics of the landscape and scenic preferences of observers.

Some examples of assessments of landscape scenic beauty are mentioned below.

Perez (2002) ascertained the landscape preferences for rambling of tourists in Spanish rural areas using a pairwise comparison of photographs. Respondents were asked to identify which of the two photos they would find preferable for rambling and to register this preference on a form by ticking the right or left box for each pair on the interview form provided. The calculation of preference was based on points accumulated every time respondents chose a picture. Their preferences were analysed according to personal characteristics and holiday activities.

Arriaza et al. (2004) assessed the visual quality of Mediterranean rural areas in Andalusia in Southern Spain firstly by ranking the agricultural landscapes on the basis of a survey of public preferences and secondly by weighing the contribution of the elements and attributes contained in the picture to perceptions of visual quality of the landscape via regression analysis. The photos used in the survey included man-made elements, positive and negative, agricultural fields, mainly of cereals and olive trees, and a natural park.

Tahvanainen et al. (1996) evaluated the scenic beauty of different rural Finnish landscapes, which were produced from nine original pictures by image-capture technology, with the open landscape becoming gradually afforested in three stages. The visual impacts of gradual afforestation were evaluated by three different interest groups of people from the study area (private non-industrial forest and agricultural landowners, potential recreationists and professionals of land use management) using two different techniques, the pairwise comparison as applied in AHP developed by Saaty (1980) and rating method using a scale from 1 to 10.

The impacts of the General Agri-Environmental Protection Scheme (GAEPS) and Supplementary Protection Scheme (SPS) on scenic beauty of Finish agricultural landscapes were evaluated using the method of Alho et al. (2001) for the scenic beauty assessment (Tahvanainen et al., 2002). The study material consisted of 3 original landscapes, which were modified to depict different management alternatives by computer-aided image-capture technology (Adobe Photoshop). The scenic beauty of the landscapes was evaluated through pairwise comparisons, and also some socio-demographic variables of the respondents were reported. In addition, a five-point Likert-scale was used to examine the attitudes of the respondents towards statements concerning environmental values, water protection, widening of buffer strips and subsidies to agriculture.

The main aim of the research of Tyrväinen et al. (2003) was to study whether aesthetic and ecological values can be combined in the management of urban forests in Helsinki, Finland. Respondents were asked to evaluate a set of photographs that were produced through digital photo manipulation (Adobe Photoshop) representing four distinctly different types of management options (e.g. no management, management of understorey and bush layer, thinning and leaving decayed trees and dead snags). The forest management alternatives were assessed using a pairwise comparison method as introduced by Alho et al. (2001). In the regression model, the preference of the landscapes (combination of landscape location and treatment) was explained with the socio-demographic characteristics of the respondents as explanatory variables. The statistical analysis was aimed at revealing how different management options were ranked, and how a given respondent's background characteristics influenced his/her evaluation.

Howley et al. (2012) analysed respondents' ratings of a variety of landscape images representing traditional extensive and more modern intensive farming landscapes of Ireland. Using a multivariate regression analysis, the relationship between demographic characteristics and people's visual preferences for traditional farming landscapes was also examined.

In the study of Ode et al. (2009), the relationship between landscape preference and three landscape indicators of naturalness (level of succession, number of woodland patches and shape index of edges) was explored using computer-generated visualisations of a hypothetical landscape containing pasture and broadleaved woodland. The respondents were asked to rate individual images using a score of between 1 and 5, based upon how much they liked the landscape. Moreover, the contribution of respondent characteristics to the formation of preference was explored using a regression analysis.

Focusing now in the Greek case study, the AE action, as mentioned before, aims to protect landscape and in particular to maintain the traditional pruning practices in vineyards of Santorini. Given that the particular traditional farming practices result in a unique agricultural landscape, this AE action contributes to improving the amenity values of the local rural environment. Amenity values are related to the aesthetic coherence as well as natural and cultural attributes of Santorini's vineyard landscape. We assume that changes in land cover, e.g. changes from the traditional pruning system to more intensive farming practices or even to some other land cover type, will decrease, by definition (since it is the main objective of the action under evaluation), the amenity services offered by the specific vineyard landscape. Thus, as a general rule we could say that if more changes appeared in the characteristic vineyards, the visual amenity will diminish and the landscape will be less attractive.

Subsequently, in order to assess the aesthetic values of Santorini's vineyards, we define the visual amenity indicator as an additional programme specific indicator that based on the results of the 'land cover change' indicator. Considering that different changes occurred in the vineyards of Santorini, the estimation of the indicator 'visual amenity' is categorised into three levels of an ordinal scale, high, medium and low. High level is considered when the pruning system is maintained, medium is when land cover (vineyard) is maintained, low is when the land cover is changed.

However, it is important to emphasize that our intention is not to conduct a public preference survey on how people value the vineyard landscape of Santorini, since the AE action explicitly state that vines pruned with the traditional techniques of 'giristi or kouloura' deliver higher visual amenity. Our case study testing aims to explore how the observed changes in vineyard landscape may affect the amenity values provided by the vineyards landscape of Northern Santorini.

Step 1.3. Defining units of analysis for micro and macro level evaluations

The use of spatially explicit data, i.e. land cover data and IACS georeferenced information, allow us to conduct a micro and macro level analysis. As in the case of land cover change indicator, the land parcel is considered the unit of analysis for the micro level analysis, and the landscape unit for the macro-level analysis.

Data Sources

Available land cover data of Northern Santorini in 2003 and 2012 drawn from the respective GE images.

The IACS geo-referenced data for field parcels participating in the AE action for the maintenance of pruning systems and also the special measures in favour of the Small Aegean Islands for the year 2011 are available (see above Data Source for land cover change indicator).

Step 1.4 Conceptual decisions on counterfactual micro and / or macro level evaluations

Given that there is a causal link established between Santorini's traditional vineyards and its contribution to making this landscape attractive for people, we assume that the alteration of iconic vineyards resulting from the characteristic pruning systems will deteriorate its distinctive landscape and consequently people's amenity will be affected negatively. Visual amenity will be estimated in accordance to three levels of an ordinal scale, high-medium-low. High level is considered when the pruning system is maintained, medium is when land cover (vineyard) is maintained, low is when the land cover is changed.

The use of spatially explicit data, i.e. land cover data and IACS georeferenced information, allow us to conduct a micro and macro level analysis.

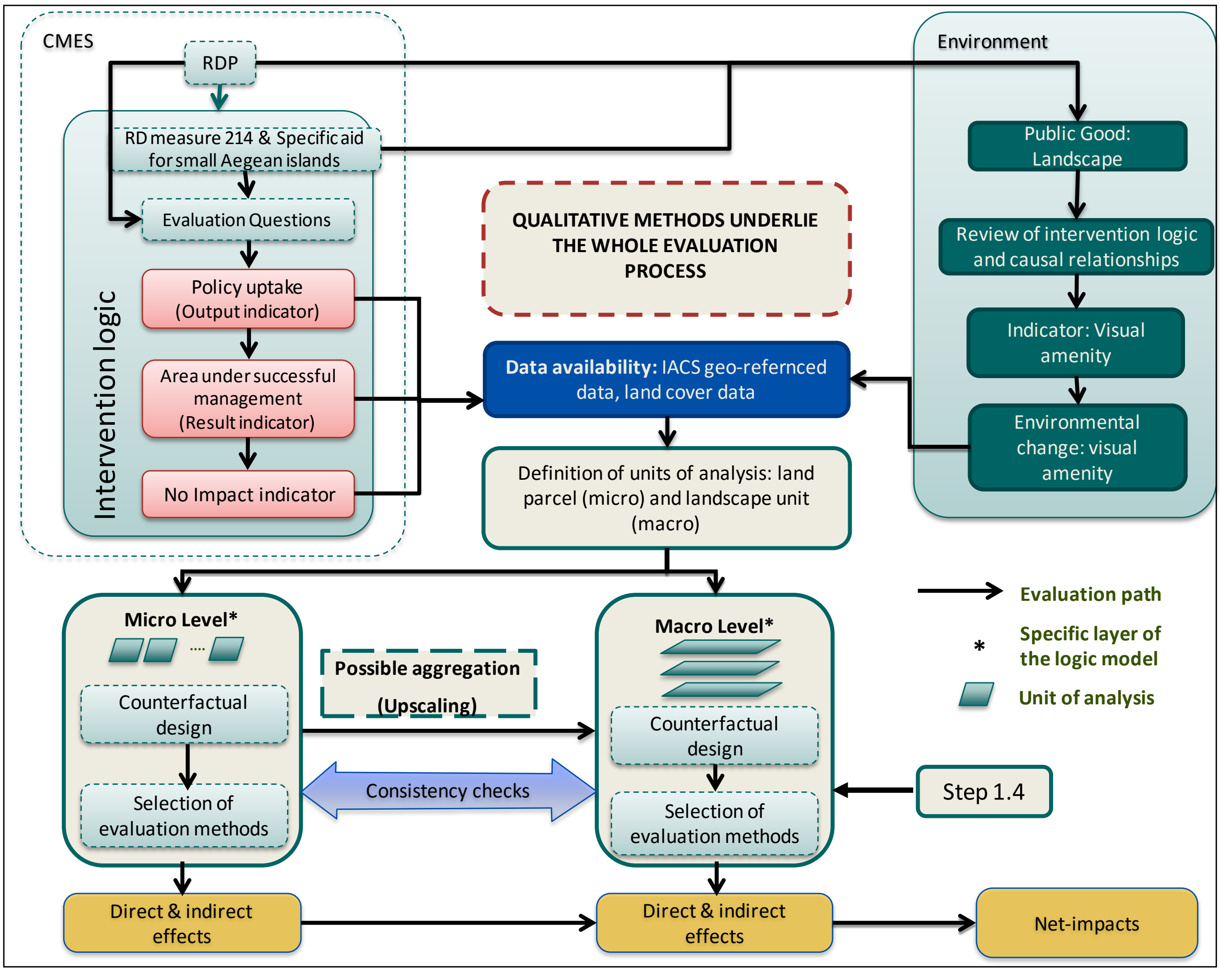

Figure 50: **Step 1.4** designing counterfactual micro and /or macro level estimations of the visual amenity indicator

Amenity values offered by the traditional vineyards of Santorini are compared among three comparison groups, i.e. field parcels participating in the AE action for the maintenance of pruning systems, field parcels participating in the AE action for the maintenance of systems alongside the special measure for continued vine cultivation, and non-participants (Table 9).

Table 6 Comparison groups for the counterfactual of visual amenity indicator

1st comparison group	2nd comparison group	3rd comparison group
Participants of AE action for the maintenance of pruning and practices in vineyards on the island of Santorini	Participants of AE action for the maintenance of pruning and practices in vineyards on the island of Santorini, and the special measures for the maintenance of traditional vines	Non-participants in either measure

Workflow and description of the counterfactual design (Steps 2.1, 2.2 and 2.3)

Counterfactual design at micro level

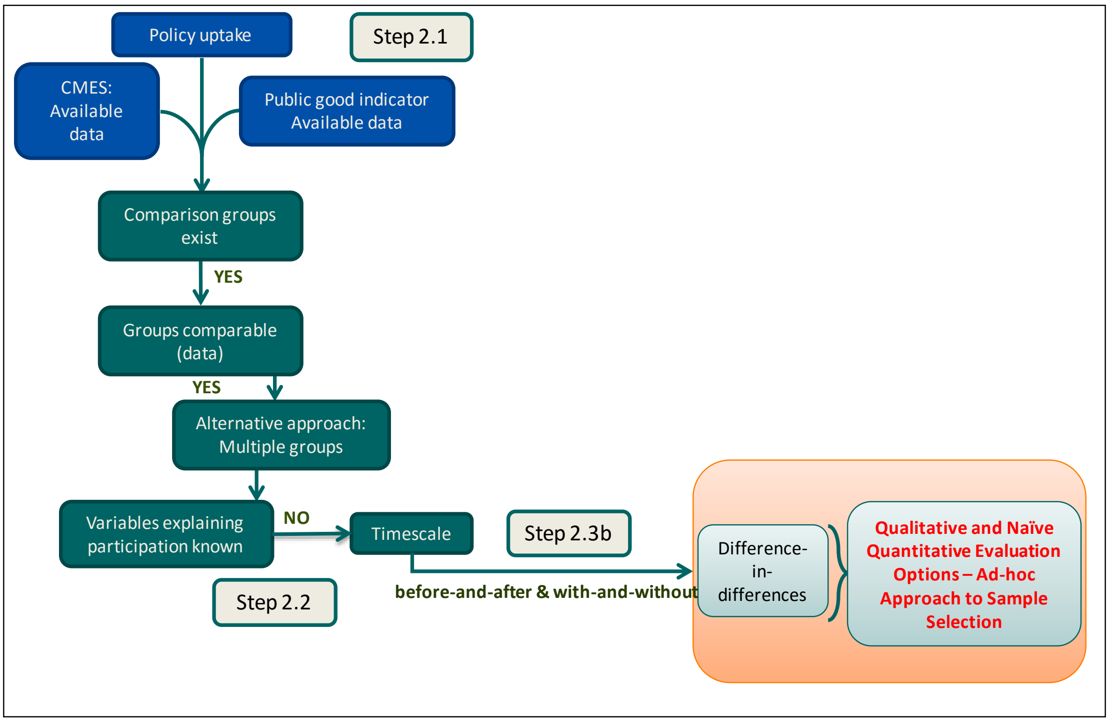

Figure 51: **Steps 2.1, 2.2 and 2.3** Counterfactual design at micro level of the visual amenity indicator

Step 2.1 Input to the counterfactual logic model

IACS georeferenced data and land cover data for participants and non-participants will be used in order to test the visual amenity indicator (see Figure 51).

Step 2.2 Defining Comparison groups

Data sets, IACS georeferenced data and land cover data, are both spatially explicit data and at the same scale and resolution, thus there is no need to include scaling mechanisms. Therefore, available data allow us to construct comparable comparison groups. There are three comparison groups, land parcels under the AE measure for the maintenance of traditional pruning practices, land parcels under the aforementioned measure and the special measure for the conservation of vineyards and land parcels under no measure. Particularly the latter group will determine our counterfactual scenario, the state of the environment without policy interventions.

Step 2.3 Selecting Counterfactual-based evaluation options

Except for the eligibility criterion for both policy measures –to be vine growers–, data does not include other variables that could statistically explain participation. Moreover, our data sets cover the time period before and after the application of the AE action –the AE action implemented since 2005 and land cover data available for the years 2003 and 2012– thus a naive Difference in Differences counterfactual approach will be used.

The amenity values based on their observed land cover changes from 2003 to 2012 will be compared among the following groups:

- land parcels participating in the AE maintenance of pruning system,
- land parcels participating in the maintenance of pruning system alongside the maintenance of traditional vine cultivation, and
- land parcels participating in neither schemes.

Visual amenity estimation is categorised into three levels, high-medium-low. High level is considered when the pruning system is maintained, medium when land cover (vineyard) is maintained, low when the land cover is changed.

Comparing the amenity values among the different groups will allow us to estimate how changes away from the traditional pruning systems may affect the attractiveness of Santorini's landscape.

Counterfactual design at macro level

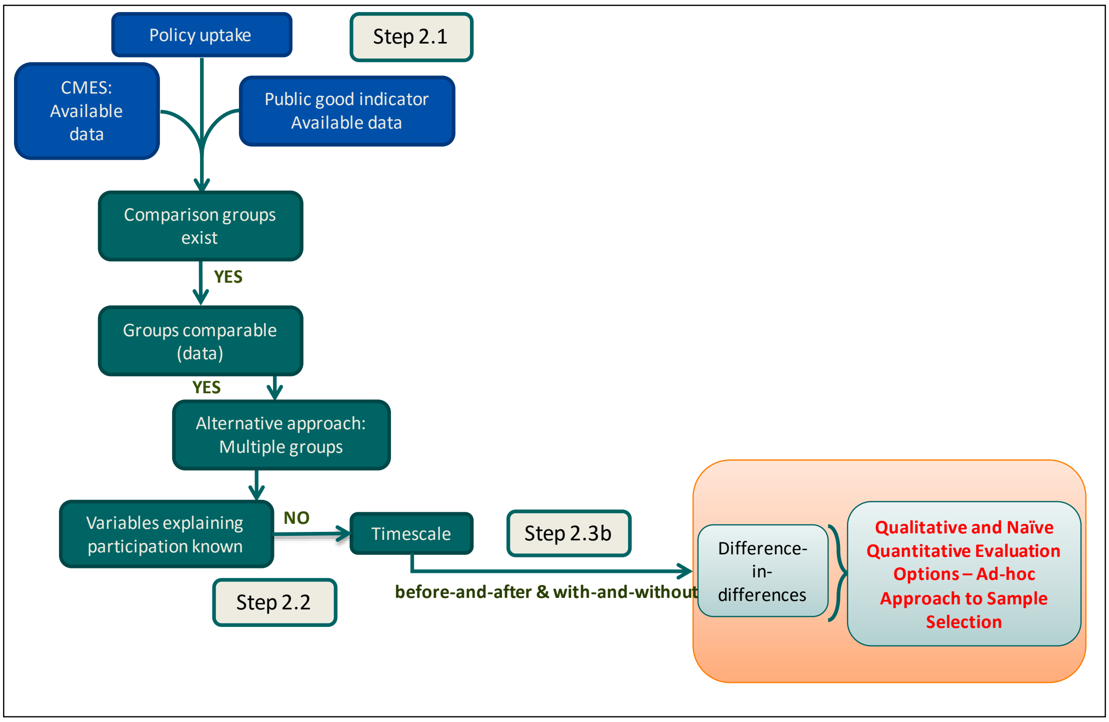

Figure 52: **Steps 2.1, 2.2 and 2.3** Counterfactual design at macro level of the visual amenity indicator

Step 2.1 Input to the counterfactual logic model

Land parcel level data will be aggregated, and analysed at landscape level. Aggregation would be the only upscaling approach to be used to link between the micro and macro level. Its consistency is attempted through the use of spatially explicit information (see Figure 52).

Step 2.2 Defining Comparison groups

IACS georeferenced data and land cover data are spatially explicit. There are three comparison groups, land parcels under the AE measure for the maintenance of traditional pruning practices, land parcels under the

AE action as well as special measure for the maintenance of vine cultivation and land parcels under no measure.

Step 2.3 Selecting Counterfactual-based evaluation options

Vine growers on the island of Santorini are eligible for participating in both policy measures, however other variables explaining participation are unknown. Further, data sets cover the time period before and after the application of the AE action. Therefore a naive Difference in Differences counterfactual approach will be used. Thus the loss of amenity values for each comparison group within the period 2003 and 2012 will be estimated.

Comparing the amenity values among the different groups will allow us to assess how these changes away from the traditional pruning systems impact on the attractiveness of Santorini's landscape.

Steps of the micro level evaluation: including contribution of the approach to assessing net impacts (Steps 3.1, 3.2 and 3.3)

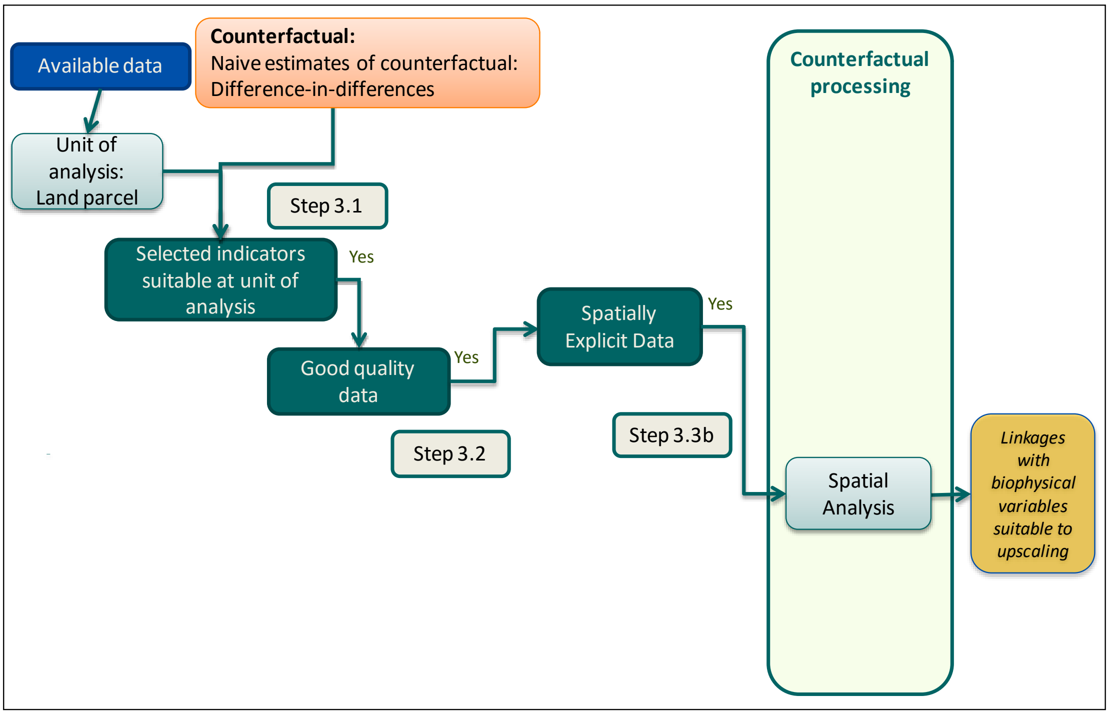

Figure 53: Layer of the micro logic model of the visual amenity indicator (Naive estimates of counterfactual)

Step 3.1. Definition of unit of analysis and the consistency of indicators

IACS georeferenced data and land cover data are available at the land parcel level, while the selected indicator visual amenity is also suitable at this micro level (see Figure 53).

Step 3.2. Assessment of data quality

IACS georeferenced data and land cover data are spatially explicit and of good quality. However information is not linked to the farm level, thus analysis doesn't allow for a robust impact assessment of policy measures.

Step 3.3b Application of naive estimates of counterfactual

Naive estimates of counterfactual will be used, since data prior and after measure implementation are available, without using particularly complex modelling approaches.

The estimation of visual amenity indicator is categorised into three levels, high-medium-low. High level is considered when the pruning system is maintained, medium when land cover (vineyard) is maintained, low when the land cover is changed. The allocation of the visual amenity level –high, medium, low– to each comparison group is based on the observed changes in vineyards' area in the timeframe 2003-2012. The frequency of occurrence and the percentage frequency of each amenity level within each comparison group are calculated. The percentage frequency distribution will be compared among the different comparison groups.

In cases where data is also linked to the farm level, it is also possible to estimate the indirect effects, such as deadweight effects, by checking changes observed in the land parcels participated in maintenance of pruning practices that would have occurred even without the applied measure.

Steps of the macro level methods: including contribution of the approach to assessing net impacts (Steps 4.1, 4.2 and 4.3).

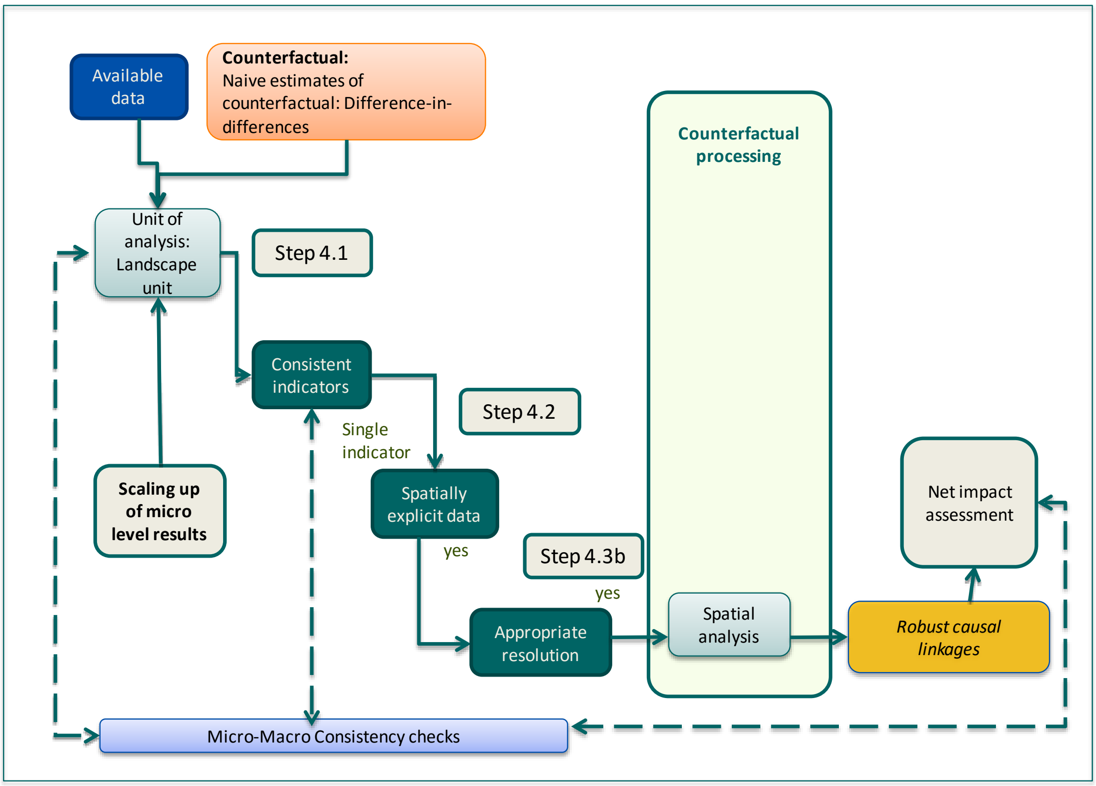

Figure 54: Layer of the macro logic model of the visual amenity indicator (Naive estimates of counterfactual)

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

Land parcel level data will be aggregated and analysed at landscape level. Aggregation would be the only upscaling approach to be used to link between the micro and macro level. Its consistency is attempted through the use of spatially explicit information (see Figure 54).

Step 4.2 Creation of consistent (spatial) data

Land parcel level data will be aggregated and analysed at landscape level. We use a single consistent indicator which is based on spatially explicit data. Moreover, the IACS georeferenced data and land-cover data, are at the same scale and resolution, thus there is no need to include scaling mechanisms.

Step 4.3 4.3b Application of naive estimates of counterfactual

Since comparison groups are comparable and there is timeseries but variables explaining participation are unknown, visual amenity as a single consistent indicator will be used for the difference in differences assessment. Given that data with geo-referenced information is accessible, this spatially explicit approach is able to improve the causality linkages. The amenity values based on their observed land cover changes from 2003 to 2012 will be compared among the comparison groups.

Steps of micro-macro linkages

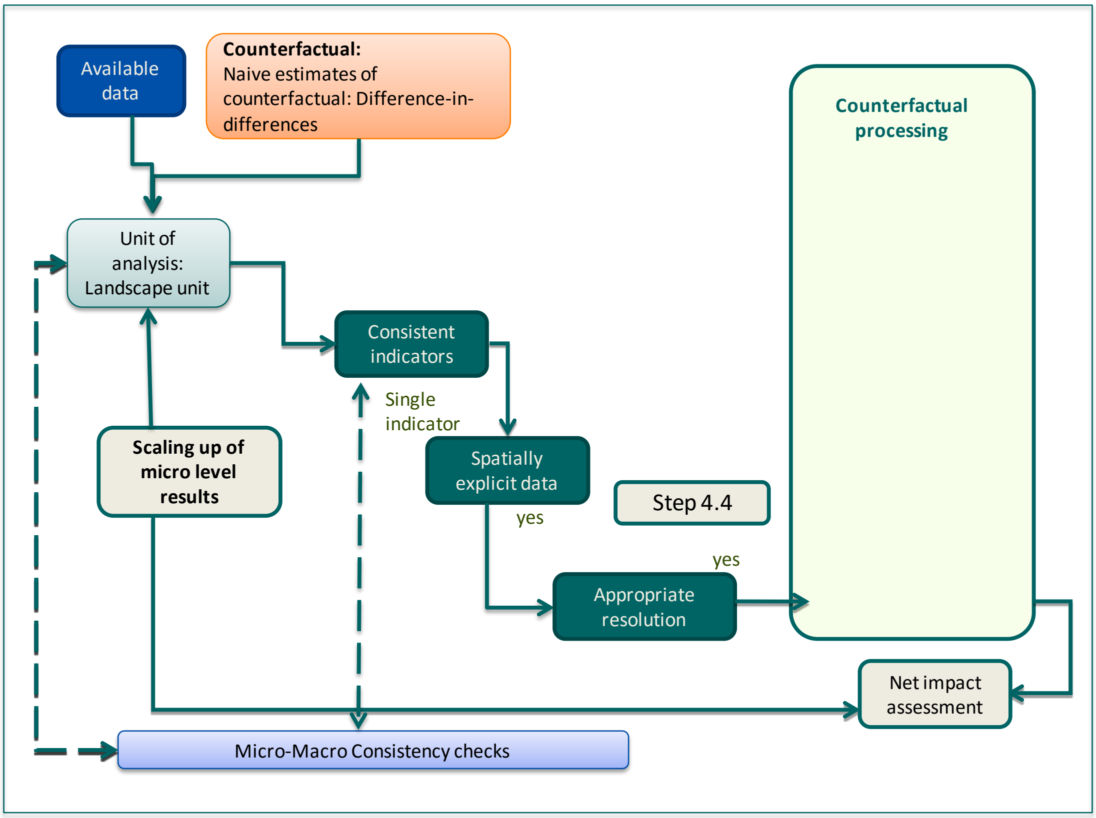

Figure 55: **Step 4.4** Micro-macro aggregation and validation of the visual amenity indicator

Step 4.4 Micro-macro aggregation and validation

Land parcel level results form the basis of the macro level assessment. Results and micro level data from land parcel level will be up-scaled and aggregated to the landscape level.

We would attempt to estimate the indirect effects, such as deadweight effects, by checking change/maintenance observed in the land parcels participated in the action for the maintenance of pruning practices that have occurred with no intervention, but lack of data at farm level inhibited such a comparison.

5 Factsheets for tested indicators and methods

The fact sheets are a final outcome presenting a short summary of the main characteristics of the indicators and methods tested in the ENVIEVAL project. 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.

5.1 Climate Stability

5.1.1 Method/Indicator: Carbon footprint

1. <u>Definition / description of the method, including:</u>

***Environmental public good*:** Climate stability

***Micro or macro-level application*:** Carbon Footprint (CF) is a well-established method to estimate carbon emission from functional units having different structural and management characteristics.

***Type of method/indicator*:** Carbon Footprint (CF) can be considered as a method to quantify GHG net emissions as well as an indicator that measure these emissions. CF has been developed in the more general setting of 'ecological footprint' (EF) proposed by Rees (1992) for measuring the human 'load' considering the human carrying capacity as the maximum persistently-supportable load. EF could be considered as a composite indicator using either a common unit of measurement (e.g. the amount of productive land and sea area necessary to supply human population consumption) or an a-dimensional value system (irrespective of the measurement unit) such as the Agri-environmental Footprint Index proposed to evaluate agri-environment schemes (Purvis et al., 2009). The CF has been developed independently, in a modified hybrid form that derives only its name from EF, but conceptually is a global warming potential indicator developed through a specific method (Pandey and Agrawal, 2014). The carbon footprint approach allows us to measure the quantity of greenhouse gases (GHGs) expressed in terms of CO2 equivalent that is directly and indirectly caused by an activity or is accumulated over the life stages of a product and emitted into the atmosphere by an individual, organisation, process, product or event from within a specified boundary. The estimation of GHG footprint can be carried out by a process-based Life Cycle Assessment (LCA) where multiple environmental impact categories are assessed from cradle to grave. A very important element in the CF assessment is the functional unit considered and its system boundaries defined in temporal and physical terms that generally depend on the subject and the policy question (Minx et al., 2009). The functional unit could be the farm and/or the single productive process and its CF is the climate impact under a specified metric that considers all relevant emission sources, sinks, and storage in both consumption and production (Peters, 2010). The CF approach focuses on emission drivers, taking into account the indirect effects of farming practice changes on other sectors, e.g. on energy sector (changes on fuel consumption) or industry (changes on fertiliser and pesticide use).

***Reflected RDP priority and focus area*:**

- Priority 5 of the RD programmes: Promoting resource efficiency and supporting the shift toward a low-carbon and climate-resilient economy in agriculture, food and forestry sectors.
- Focus area 5D: Reducing greenhouse gas and ammonia emissions from agriculture; Focus area 5E Fostering carbon conservation and sequestration in agriculture and forestry.

***Unit of measurement*:** CO_2 equivalent that describes, for a given mixture and amount of greenhouse gas, the amount of CO_2 that would have the same global warming potential (GWP).

2. Micro or macro-level application:

Emission drivers need to be considered at several scales and in different contexts, using different functional units and methods (Peters, 2010). The scale of application is at farm level and/or at process level, and the regional level can be derived through consistent upscaling of available representative farm-level or process-level data. In this case, the issue of double counting of holdings in the middle of the supply chain is not relevant due to the fact that the functional unit only refers to farms and not to other suppliers along the chain.

3. General requirements

Data requirements**:** Land-cover data (UAA area and crops from FADN, FSS-Agricultural Census; IACS; LPIS); general farm data (FADN, Agricultural Census); input-use data at farm level and single process level for crops and animals (FADN); production of fuel, electricity, machinery, fertiliser, pesticide, and plastic used in the production processes and emissions during the production of any replacement animals (FADN and scientific literature) and data on soil conditions. The existing databases (e.g. FADN, FSS) are usually not sufficiently detailed in terms of information needed to create robust estimation and ad-hoc surveys are generally requested to provide additional information.

Skill requirements: biophysical approaches (es. LCA, input data interpretation, etc.), statistical analysis, bibliographical review skill.

4. Consideration of counterfactuals

Carbon footprint methods provide input for counterfactual approaches. Where there is sufficient data availability (i.e. samples with more than 30-50 observations for each group), quantitative methods linked to quasi-experimental design could be applied. For example, Propensity score matching matches participants to similar non-participants for statistical analysis. The use of different data sources (FADN, FSS and IACS) should also guarantee the analysis in the temporal dimension. In those cases, the control group design depends on availability of data required at farm level (or cadastral parcel in case of process level) for participants and non-participants. Where there is weak data availability, naïve estimate of counterfactuals (with-vs-without approach and before-and-after) could be used.

5. Context of the case study testing

Case study area**:** Emilia Romagna Region of Italy

Policy context**:** In the past RDP the main measures of reference were: Agri-environment sub-measures for climate change (214/A, 214/B and 214/E); Support for non-productive investments (216) and Increased renewable energy production (221).

Data used**:** Ad-hoc survey for crops and livestock, where primary monitoring data related to land use and input use were collected by the evaluator. The IACS database will be used to distinguish participants to RDP programme from non-participants.

Evaluation approach tested**:** The analysis of the GHG emission at process level has been carried out comparing different farming systems (organic and integrated vs. conventional ones), which have proven to be the best functional units to apply the CF methods. At farm level, the calculation of direct and indirect GHG emission can be done with the JRC Carbon Calculator (Bochu and

Metayer, 2013), a user-friendly open-source tool designed to assess the life cycle of GHG emissions from different types of farming systems. The information contained in the FADN sample could be a good starting point, although additional information about farming practices is needed. CF allows for the creation of comparison groups of before-and-after as well as with-and-without participants. However, when the number of observations is insufficient for an elaborate statistics-based evaluation, a naïve group comparison counterfactual approach can be used, possibly assuring that sample selection has been minimised through expert knowledge to create similar comparable groups.

Table 7 Strengths and weaknesses of the method (based on SWOT)

Evaluation challenges (relevant for methods)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	The use of site-specific data (FADN and direct surveys) allows taking account of specific local situations	Difficulty in considering matters that are strictly related to farm management (farming practices)	CF allows the measurement of RDP impacts at different levels
Timing of environmental impacts captured	Use of FADN data provides annual data by farms without the needs of specific environmental impact monitoring	Additional ad-hoc surveys are only done occasionally (e.g. RDP evaluation)	CF allows periodic assessment of the environmental performance of RDP measures
Establishment of robust causal relationships	CF allows a direct judgment of GHG performance of the single productive (functional) unit and is based on a theoretically sound basis	GHG emissions are calculated based on average CO_2 emission coefficients applied to individual farming systems, making them often unrepresentative in a local context. Data processing is quite time-intensive	An estimation of the overall effects of GHG emissions (including offsite emissions) is possible with the LCA approach
Assessment of net-impacts	The estimation of direct and indirect effects can be done for each of the control groups	FADN data allow only use of naïve methods for low number of observations for control groups	Process-level analysis could reduce the need for primary data monitoring
Establishment of consistent micro-macro linkages	The method allows inference from the micro to the macro level (regional) by upscaling	Good upscaling depends on sample territorial representativeness and this is not always possible with process-based or FADN samples	A statistically-sound representativeness of ad-hoc surveys can ease the upscaling from farming system unit to regional level.
Appropriateness of indicator(s)	CF synthesises well the complexity of environmental relationships behind the GHG emissions		
Unambiguous and understandable results	The results can be easily communicated to target groups (managing authorities, policy makers and farmers)	The inner workings of calculating the CF are quite complex	

6. Recommended application

The estimation of GHG emissions and sinks is generally a time-intensive procedure which requires adequate expertise and could undermine its use due to very high costs for ad hoc monitoring surveys and data processing. Options to overcome this limitation can come from: a) the realisation of multi-purpose surveys on farming practices useful for more than one impact indicator; b) the design of software tools that could be a worthwhile investment beyond the monitoring and evaluation phase, increasing the use of these tools in the farmer decision-making process to improve performance efficiency.

Current existing databases such as FADN and IACS cannot provide all the types of required data for the carbon footprint calculation, lacking information on farming practices. Moreover, some problems may occur when the reference databases are not statistically representative about the quality and quantity of inputs purchased and used by farmers and the type of implementation of farming practices. Furthermore, the more are the variables to be considered in the production systems, the more CF analysis is complex (e.g., mixed farms compared to mono-cultural farming systems).

The lack of sector- and region - specific emission factors for important agricultural inputs add to the uncertainty. The standard method must address how to deal with alternative scenarios and land use changes.

5.1.2 Method: Sector models - DREMFIA

1. Definition / description of the method, including:

Type of method: Multi-region dynamic partial equilibrium model

Environmental public good: Climate stability

Micro or macro-level application: DREMFIA (Dynamic multi-REgional sector Model for FInnish Agriculture) has been developed over the years to simulate agricultural production and markets in Finland from 1995 to 2020. The macro-level model is based on spatial (regional) price equilibrium assuming competitive markets with basic profit and utility maximising conditions for producers and consumers alike. Each region specialises in products and production lines that yield the greatest relative profitability, taking into account the profitability of production in other regions and consumer demand. Use of different production resources, including farmland, in different regions is optimised in order to maximise sectoral welfare, taking into account differences in resource quality, technology, costs of production inputs and transportation costs. The DREMFIA model consists of two main parts: (1) a technology diffusion model that determines sector-level investments in different production technologies; (2) an optimisation routine simulating annual production decisions (within the limits of fixed factors) and price changes, i.e. supply and demand reactions, by maximising producer and consumer surplus subject to regional product balance and resource (land and capital) constraints. In part (1), production activities include a number of different animals, hectares under different crops and set-aside, feed diet composition, chemical and manure fertiliser use and the resulting crop yield level. Products and intermediate products may be transported between the regions at certain transportation costs. In part (2), technical change and investments, which imply evolution of farm-size distribution and production capital in different regions, are modelled as a process of technology diffusion. In a dynamic recursive model, parts (1) and (2) interact each year so that prices from the market-simulating optimisation model enter the technology diffusion model, representing sector-level investments in each region, and changes in animal production capacities of different techniques enter the market model in the following year. Foreign trade activities are included in DREMFIA with imported and domestic products considered as imperfect substitutes. Climate effects on the environment are an archetypical example of global pollutant effects where a single small emitter's effect cannot be quantified by other than pressure indicators. The effects are therefore studied on a regional (macro) rather than farm (micro) scale.

2. General requirements

***Data requirements*:** As DREMFIA is an up-and-running sectoral model, all the data required for analysis are collected on a continual basis. Further, the complexity of the model requires multiple data sources, partly from official statistics and partly from other sources.

***Skill requirements*:** Building up such a sector model as DREMFIA is time consuming, demanding and requires advantage skills. Also using the model needs trained personnel.

3. Consideration of counterfactuals

Sector models can be used to model multiple counterfactuals. Essentially these models can cover the lack of data-based comparison groups. The challenge of the evaluator is to determine: what is the relevant counterfactual to be considered in impact evaluation. For example, the removal of AE payments without any compensation to farmers through other measures may not be a viable political scenario for the counterfactual.

4. Context of the case study testing

DREMFIA results can be disaggregated from national to regional level. Regional effects are calculated for four main areas and 17 sub-areas. Farm-specific effects are not possible to assess with the model, but are also not as relevant due to the global nature of the assessed pressure indicator (i.e. CO_2 equivalent emissions).

As the majority of Finnish AEM sub-measures do not specifically target GHG emissions (except for the special measure for long-term grass/hay growing peat fields which has relatively few participants), we use the grouping of the AEM (214) and LFA (211, 212) for policy analysis (MMM, 2007). These two schemes affect the overall land use and production intensity – major contributors to agricultural CO_2 emissions – which the model essentially captures.

DREMFIA can construct counterfactuals without real-world comparison groups using a wide variety of data describing both domestic and international market conditions. Exogenously-determined EU prices influence domestic prices, but domestic prices may be different from EU prices. Four main areas are included in the model: Southern Finland, Central Finland, Ostrobothnia (the western part of Finland) and Northern Finland. Production in these areas is further divided into sub-regions on the basis of the support areas. In total, there are 17 different production regions. This allows a regionally disaggregated, exact description of policy measures and production technology.

DREMFIA uses multiple data sources to simulate the agricultural markets in Finland. The simulation model uses annual-level statistics collected between years 1995-2012 and also has its own collection of data. The used data sources are best represented as a list due to their number:

Data from official statistics used in simulation:

- Prices of agricultural inputs, commodities and dairy products
- Consumption of agricultural commodities and dairy products
- Imports and exports of agricultural commodities and dairy products
- Use of crops as fodder at farms and in fodder industry
- Production yields per hectare and per animal

Data from official statistics used in model validation:

- Agricultural total calculations on the value of different inputs in agriculture (similar to EEA)
- Land use under different crops and number of animals at different regions and in the whole country
- Farm structure statistics (FSS) – distribution of dairy cows in different farm size categories is endogenous in the model

Data partially available in official statistics:

- Agricultural payments (according to support regions and specific rules and definitions)
- CAP pillar 1 payments
- LFA-payments
- Agri-environmental programme
- National subsidies
- Investment subsidies – specific to various kind of investments in livestock and crop production

Other data:

- Use of inputs in agricultural production - per ha, per head per year
- Mainly from activity-based cost models maintained and published by national agricultural extension services (www.proagria.fi)
- Partly from FADN – activity based unit cost calculations
- Use of different feed stuffs per animal, from dairy farm recording system, and other livestock specific data systems of agricultural extension services (www.proagria.fi)

Other knowledge used in simulation:

- Specific needs of energy and protein content as well as roughage needs of different animals – Luke (Natural Research Insitute Finland) feeding norms
- Nutrient contents of manure of different livestock
- Luke internal calculations maintained in animal nutrition research and/or specific tables retrieved and summarised in different research projects
- Nitrogen response function parameters
- Milk yield response function parameters
- Other technical parameters related to use of inputs per ha and head

Evaluation approach tested**:** At first we identified the grouping of the AEM (214) and LFA (211, 212) as the relevant measures for policy analysis. The evaluation question is essentially how much the agri-environmental measures have contributed to greenhouse gas emissions. Then we noted that the environmental change in terms of climate change is not a feasible measure, as long-term trend change is not observable within the evaluation period and is very hard to quantify per farm or region. Thus the indicators used in the case study employ CO_2 equivalent measures both with and without land-cover changes (LULUCF). As DREMFIA is an up-and-running sectoral model, all the data required for analysis are already collected. The majority of Finnish agricultural producers are long-term participants in agri-environmental measures, making the construction of direct comparison groups (with-and-without or before-and-after within the evaluation period) impossible. Thus the counterfactual methodology requires the use of methods that can cover the lack of data-based comparison groups. DREMFIA was deemed to be more than sufficient for this evaluation case study, thus leading to Evaluation Options without Comparison Groups in the logic model.

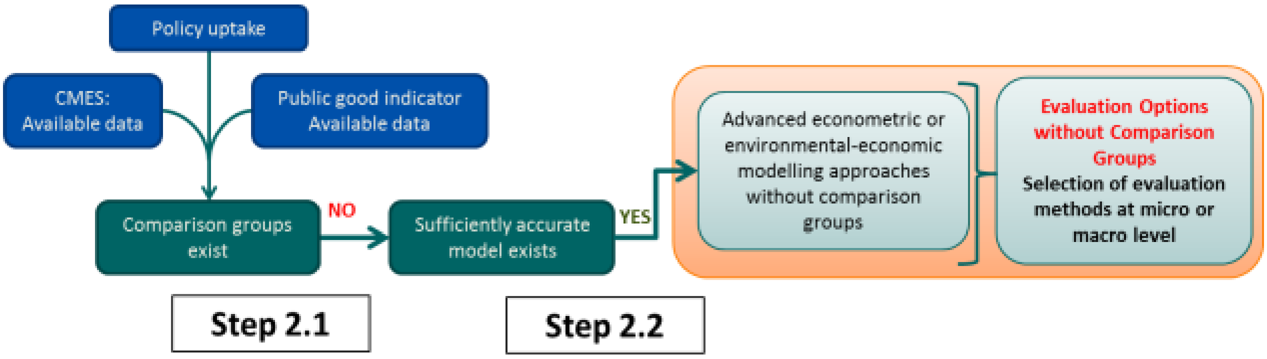

Figure 56: Evaluation steps of counterfactual logic model

Table 8 Strengths and weaknesses of the method (based on SWOT)

Evaluation challenges	Strengths	Weaknesses	Key contribution to evaluation benefits
Compatibility with local environmental and farm structural characteristics		Regional level modelling does not consider local environmental characteristics.	
Timing of environmental impacts captured	The dynamic optimisation procedure has very high temporal resolution, and can show when effects are happening.		The modelling allows dynamic impact assessment also on the post-evaluation period.
Establishment of robust causal relationships	The model is well documented and transfer functions explaining environmental effects are based on relevant scientific literature. Policies are directly modelled with all their requirements, providing excellent grounds to understand how policy mixes work together (assuming profit maximisation).		
Establishment of consistent micro-macro linkages		The model does not incorporate farm level optimisation, rather a regionally representative farm, thus losing spatial resolution. However,	

Evaluation challenges	Strengths	Weaknesses	Key contribution to evaluation benefits
		increasing the spatial resolution would come at a great computational and design cost, and might not be feasible	
Assessment of net-impacts	The model can capture substitution effects.		The model allows for comparative repetition of the impact assessment with new data.
Appropriateness of method to capture complexity of environmental relationships	The complexity of the model allows for testing numerous counterfactuals and assumptions to see if the evaluated policy had an impact. For climate in particular, the model provides clear results on impacts. Other measurable environmental indicators include diffuse water pollution indicators, providing a chance to examine joint effects of policy.		
Unambiguous and understandable results	The results are quantitative and take into account the complex structure of the agricultural production and environmental effects in Finland. The results are highly useable in policy work, providing also a chance for ex-ante recommendations.		The model allows for a number of counterfactuals, thus allowing the policy makers to refine paths of development. The model also incorporates other public goods in the analysis, providing a chance for a more holistic impact assessment.

5. Recommended application

Sector models can cover a lack of comparison groups and are flexible in handling a number of counterfactual scenarios. Sector models can also deal with displacement and substitution effects. However, construction, updating and using the models require constant funding and persons trained in their use.

5.2 Soil Quality

5.2.1 Method: InVEST model

1. **Definition / description of the method, including:**

Type of method: Biophysical model

Environmental public good: Soil quality

Micro or macro level application: InVEST stands for Integrated Valuation of Ecosystem Services and Trade-offs (Sharpe et al., 2015). It is a suite of spatially-explicit models for a number of distinct ecosystem services, and enables an assessment of quantified trade-offs associated with alternative management choices. InVEST has a flexible spatial resolution which means that it can address questions at local, regional and global scales depending on the input data quality.

2. **General requirements**

Data requirements: the InVEST data requirements are model dependent. The *carbon storage and sequestration* model requires land use/land cover (LULC) data and carbon in soil, in biomass (above and below ground) and in dead organic matter, which will calculate total carbon stock (Mg/pixel) and carbon sequestration rates (Mg/pixel/yr). The *sediment retention* model requires LULC data, digital elevation model, rainfall erosivity, soil erodibility, crop factor, management factor, sediment retention efficiency by LULC, slope threshold, flow accumulation threshold and (sub)-watershed data. This model will calculate mean annual erosion (tons/watershed/yr) and mean sediment retention (tons/watershed/yr).

Skill requirements: Spatial analytical /GIS skills

3. **Consideration of counterfactuals**

The InVest model provides input for counterfactual approaches. Where there is sufficient data availability, quantitative methods linked to quasi-experimental design could be applied. For example, propensity score matching matches participants to similar non-participants for statistical analysis. The control group design depends on availability of data (e.g. IACS/LPIS geo-referenced land-use data of 2008-2013) required for beneficiaries and non-beneficiaries. Where there is weak data availability, naïve estimate of counterfactuals (with-vs-without approach and before-and-after) could be used.

4. **Context of the case study testing**

Case study area: Aberdeenshire, Scotland

Policy context: Agri-environmental and forest-environment measures include objectives for maintaining soil carbon and avoiding soil loss through erosion

Used data: The LULC data are provided by IACS/LPIS geo-referenced summarised land-use data of 2008-2013, and national land-cover data (LCM2007) are used to fill any gaps in the IACS land-use data. Measure uptake data for soil-relevant measures (214, 223 and 225) were used. Other data requirements are derived from additional local data sources (GB Ordnance Survey data, Scottish Soil data and expert knowledge).

Evaluation approach tested: The data available to parameterise the two InVEST models that were used limited the assessment to sub-catchment level summaries of the indicators. Therefore a macro level only assessment was conducted, with an assessment based on a comparison of sub-catchment with and without participation. For each year of the RDP, the model calculates the indicators based on the LULC data. The limited information about the comparison groups meant that a naive counterfactual approach was used to compare before and after for sub-catchments with and without participation based on simple mean values.

Table 9 Strengths and weaknesses of the method (based on SWOT)

Evaluation challenges	Strengths	Weaknesses	Key contribution to evaluation benefits
Compatibility with local environmental and farm structural characteristics	The model uses existing available data taking into account important crop types and soil conditions of the case study area	The model is solely based on land-use data; it only measures that which creates measurable change in land use	The impact of the AE action is estimated for an area where the soil quality assessment is constrained by the lack of observational soil data.
Timing of environmental impacts captured	The model is able to use land-use data available through IACS to model change in soil quality indicators.		
Establishment of robust causal relationships	The method is based on a well-documented theoretically-sound model linking the land use and environmental outcomes.	The obtained results were not verified with monitoring soil quality data (lack of time and suitable soil data).	The model calculated soil erosion and sediment retention in kg/ha, mean organic carbon content in tonnes/ha and total estimated organic carbon content in arable land (in megatonnes) for sub-catchments with and without participants before and after.
Establishment of consistent micro-macro linkages	Field-level land-use data are used to model the soil quality indicators; however the assessment is strictly macro level (sub-catchment)	The model is not suitable to include farm-level results due to modelling uncertainty.	Micro and macro linkages exist only through the micro-level input data to the modelling process. However, the modelling approach does allow a macro-level assessment with a naive counterfactual in the absence of good quality observational data.
Assessment of net-impacts	The estimation of direct and indirect effects needs the availability of control groups.	IACS data constraints allow only a naive DiD counterfactual approach. Data to explore in detail the changes between different comparison groups were missing.	Despite the data availability issue, the InVest modelling approach has shown its ability to inform the net-impact assessment at macro level.
Appropriateness of indicator(s) to capture complexity of environmental relationships	The model does incorporate the complex environmental and spatial relationships in its calculation of the indicators.	The obtained results were not verified with monitoring soil quality data.	The biophysical model provides results based on site-specific environmental conditions enabling an impact assessment in the absence of suitable soil monitoring data.

Evaluation challenges	Strengths	Weaknesses	Key contribution to evaluation benefits
Unambiguous and understandable results	Results are easy to communicate to laypersons.		Method provides user-friendly outcomes in the form of maps.

5. **Recommended application**

The InVEST suite of models is developed to support the decision making in relation to a range of ecosystem services. Commonly this approach is used to consider the impact of changing LULC into the future and inform decision making; however it has proven to be suitable too for an ex-post assessment of RDP impact on soil quality for circumstances with limited observational soil data. The quality of the model input data determines the level at which the results can be used and the type of comparison groups that can be designed for the assessment. For the application of an elaborate statistics-based counterfactual method, more explanatory information regarding the RDP measure uptake should be available.

5.3 Water Quality

5.3.1 Method: Biophysical model

1. <u>Definition / description of the method, including:</u>

Type of method (linking back to the classifications used in the review deliverables): Biophysical model

Environmental public good: Water quality

Micro or macro level application: Land parcel level. The availability of data at parcel level will allow the aggregation to the upper and measure level.

2. <u>General requirements</u>

Data requirements: Water use and fertilisation input use, monitoring data at farm level

Skill requirements: Spatial analytical /GIS skills

3. <u>Consideration of counterfactuals</u>

The biophysical model can be used for Qualitative and Naive Quantitative Evaluation Options and Statistics-based Evaluation Options. The quality and quantity of the data play the key role in the construction of comparison groups. Thus the availability of sufficient data on participants and non-participants before and after measure implementation determines what counterfactual approach will be used.

4. <u>Context of the case study testing</u>

Case study area: Karditsa regional department, Thessaly Plain, Greece

Policy context: AE action for the reduction of nitrate pollution caused by agriculture in NVZs

Used data: IACS geo-referenced data of 2011 for participants and non-participants including the number of hectares of supported area and type of crop, a soil map of the specific site of the plain area of Karditsa.

Evaluation approach tested (short explanation of the main logic model steps): Land parcel level and the specific site of the NVZ of Karditsa were the units of analysis for the micro and macro level, respectively. Given that IACS geo-referenced data were only available for the year 2011, two comparison groups were constructed. Thus, a naive counterfactual approach was used comparing only participants in the nitrate pollution reduction scheme and non-participants focusing on land potentially irrigated and cultivated under intensive crops.

Table 10 Strengths and weaknesses of the method (based on SWOT)

Evaluation challenges	Strengths	Weaknesses	Key contribution to evaluation benefits
Compatibility with local environmental and farm structural characteristics	The biophysical model uses existing available data taking into account important crop types, soil conditions of the case study area in relation to the applied	Actual information on fertiliser application and water use is missing.	The impact of the AE action is estimated within each soil class taking into account the different farming practices applied.

Evaluation challenges	Strengths	Weaknesses	Key contribution to evaluation benefits
	different farming practices of the AE action		
Timing of environmental impacts captured	Use of a static biophysical model that is based on existing data.	The impact of the AE action cannot be captured within the timeframe of the evaluation.	
Establishment of robust causal relationships	The method is based on a well-documented theoretically-sound model linking the farming practices and environmental outcomes.	The obtained results were not verified with monitoring water quality and quantity data (Lack of time).	The biophysical model calculated the GNB in the form of nitrogen losses per ha and the water use/ha between participants and non-participants.
Establishment of consistent micro-macro linkages	Macro-level analysis can be built on aggregated micro-level results.	Farm level which is the decision level for participation in the various schemes was missing.	Micro and macro linkages considered only in an intuitive manner. Two macro-level analyses have been used. The first was based on the assumption that each crop type is distributed with the same percentage in each soil class as in the total case study area; the second on the actual distribution across soil classes.
Assessment of net-impacts	The estimation of direct and indirect effects needs the availability of control groups.	IACS data constraints did not allow the application of a DiD counterfactual approach exploring changes between different comparison groups over time.	The biophysical model provided quantifiable results.
Appropriateness of method to capture complexity of environmental relationships	The biophysical model suggests the maximum amount of nitrogen for significant crops and soil classes as well as the rational irrigation rates for significant crops and soil classes in order to avoid groundwater overexploitation.	The obtained results were not verified with monitoring water quality and quantity data.	The biophysical model provides results in relation to the site specific environmental conditions.
Unambiguous and understandable results	Results are easy to communicate to laypersons.		Method provides user-friendly outcomes.

5. Recommended application

The applied biophysical method calculates the nitrogen fertiliser application, nitrogen loss and water use/ha for specific crops taking into account the soil texture, relief and the nitrogen balance equation. Therefore this model is recommended in cases where actual data on water quality and quantity, i.e. fertiliser and water use, are missing. The required data includes IACS/LPIS data between participants and non-participants at different points in time and a detailed soil map.

Moreover, when data with geo-referenced information is accessible, this spatially explicit approach has a great potential improving the causality linkages.

5.3.2 Indicator: Mineral nitrogen content in the soil in autumn (Nmin indicator)

1. <u>Definition / description of the indicator, including:</u>

Environmental public good: Water quality

Type of indicator: Complementary result indicator. The autumn Nmin value provides information on the amount of nitrogen in the soil that is potentially polluting the groundwater due to leaching during winter.

Reflected RDP priority and focus area:

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
- Focus area 4B: Improving water management, including fertiliser and pesticide management.

Unit of measurement: Amount of nitrogen per ha (kg N/ha) in the form of nitrate or ammonia, in soil depths between 0 and 90 cm.

Type of data required:

- Soil samples of the mineral nitrogen content of soil in autumn (monitoring data) of participating and non-participating sites with a sufficient sample size (minimum 100 samples per (sub-) measure).
- Farm structural data including information on site specific conditions (e.g. sink or source characteristics, soil type), farm management practices (type of crops, type of grassland-use, livestock density) and weather conditions.

Scale and level of application: parcel level (spatial)

2. <u>Existing data sources</u>

EU-level: Data is not available at EU level

***MS and regional level* (examples)**: In Lower Saxony, Germany, the data is collected for monitoring purposes by the managing authority. It is usually analysed at the level of the drinking water extraction areas.

Fact sheets and information available from other sources: A description of the indicators' characteristics and their application for impact assessment of AE measures was published by the monitoring organisation of Lower Saxony but is only available in German. (NLWKN, 2015c and NLWKN, 2010).

3. <u>Context of the case study testing</u>

Case study area: Lower Saxony, Germany

Policy context*:* Water protection measures to reduce diffuse pollution from agriculture are a key policy objective for agri-environmental policies.

Used data: Monitoring data of roughly 20,000 soil samples for the years 2000 to 2006 were used for micro-level analysis. For the years 2008-2012, only data aggregated at the level of drinking-water protection areas (i.e. important ground water areas) were available for the analysis.
Evaluation approach tested: Nmin results for soil samples that were collected by the managing authority for monitoring purposes were used for the analysis. Two comparison groups were used to compare sites with AEM participation with non-participants. Samples of sites with similar environmental conditions were matched with each other. One site with AE measure was compared to three sites of non-participants. A pairwise comparison and regression analysis were conducted at micro level. The analysis at macro level used the aggregated data set that was provided by the managing authority for the recent years.

Table 11 Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Typology according to the available data on environmental conditions and farm structure is used. Required information: - site specific conditions - farm management practices - weather conditions Analysis is based on measurements.	Data gaps on local conditions and farm structure can limited the application of the indicator with elaborate statistics-based approaches. Large sample size necessary. Samples stem from land parcels with different regional and temporal distribution.	Use of existing monitoring data for the evaluation. Use of a matching approach to compare similar farms.
Appropriateness of indicator(s) to capture complexity of environmental relationships	Indicator is a proxy with a strong linkage to the potential nitrate pollution of the groundwater. It provides a reasonable option to deal with time lags until impacts can be measured in ground water. Indicator is well known and used in case study area. Indicator delivers measurements of change and impact	Impact measured with a proxy might not reflect actual effects in the ground water. Indicator is not used in other regions. Comparison with other programmes is not possible.	Additional suitable impact indicator was tested

4. <u>Recommended application</u>

Autumn Nmin value can be used as a complementary result indicator for the evaluation of measures and sub-measures at parcel level. The timing of the measurement of the Nmin indicator is very close to the implementation of the AE measures. Thus, it is well suited for the annual impact assessment of water protection measures on agricultural land.
Basic sampling requirements for a robust impact assessment:

1. Suitable site-specific conditions
2. A minimum of 16 punctures per area
3. Sample taking from October to mid-November (before the leakage water formation)
4. If precipitation in autumn is high, sampling depth has to be adapted (deeper than 90 cm) (NLWKN, 2015c)

Timing of sampling can reduce risk of bias due to climatic conditions.

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 and farm structural 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.

5.3.3 Indicator: Water use/ha

1. <u>Definition / description of the indicator, including:</u>

Environmental public good: Water quality

Type of indicator: The new CMES impact indicator is water abstraction in agriculture (European Commission, 2013). This indicator refers to the volume of water which is applied to soils for irrigation purposes. Data concern water abstraction from total surface and ground water.

Reflected RDP priority and focus area:

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
- Focus area 4B: Improving water management, including fertiliser and pesticide management.

Unit of measurement: m3 of water abstracted for irrigation

Type of data required: The indicator of water use/ha is estimated by a biophysical model using the applied water per dose, a detailed soil map and type of crop. LPIS-GIS data for participants and non-participants at different points in time are needed. Since IACS-LPIS data are available at different points in time, the deadweight effects could also be estimated.

Scale and level of application: The indicator on water abstraction could be calculated at NUTS 2 level ideally (and River Basin level); an analysis at regional level is more appropriate to capture the effects and impacts of the CAP on the environment.

2. <u>Existing data sources</u>

EU-level: The Survey on Agricultural Production Methods (SAPM) provides estimates of water use for irrigation on farm level. SAPM is a unique survey carried out by Eurostat in 2010 to collect data at farm level on agri-environmental measures. Data on water abstraction for irrigation cannot yet provide a pan-EU coverage (Eurostat, Water abstraction).

MS and regional level (examples): Annual data available for the period 1970-2009 depending on availability for each MSs (In 2007, 2008, 2009 data are available for 19, 11, 10 MSs respectively, Eurostat/OECD Joint Questionnaire).

Fact sheets and information available from other sources:

- IRENA, Indicator Fact Sheet 34.3-Share of agriculture in water use
- IRENA Indicator Fact Sheet 22-Water abstraction

3. <u>Context of the case study testing</u>

Case study area: Karditsa regional department, Thessaly Plain, Greece

Policy context: AE action for the reduction of nitrate pollution caused by agriculture in Nitrate Vulnerable Zones (NVZ)

Used data: IACS geo-referenced data of 2011 for participants and non-participants including the ha of supported area and type of crop, a soil map of a specific site of the plain area of Karditsa
Evaluation approach tested: Land parcel level and the specific site of the NVZ of Karditsa were the units of analysis for the micro and macro level, respectively. Given that IACS georeferenced data for participants and non-participants were only available for the year of 2011, two comparison groups were constructed, using a naive counterfactual approach comparing only participants in the nitrate pollution reduction scheme and non-participants.

Table 12 Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Using the water use/ha in combination with a detailed soil map enables the consideration of specific environmental characteristics, such as soil texture and relief, as well as significant crops.	Missing information on farm characteristics (e.g. use of agricultural area, type of farming system, participation in RDP or other policy measures) can constraint the application of the indicator in elaborate statistics-based evaluations.	The impact of the AE action is estimated within each soil class.
Appropriateness of indicator(s) to capture complexity of environmental relationships	The indicator is calculated by a biophysical model that proposes rational irrigation rates for significant crops and soil classes in order to avoid groundwater overexploitation.	Actual data on water used for irrigation purposes are missing.	The impact of the AE action takes into account the different farming practices applied.

4. <u>Recommended application</u>

The specific indicator may provide useful information on agri-environmental schemes that promote the sustainable management of water resources. The water use/ha indicator was estimated by a biophysical model and analysed in relation to the different farming practices of the AE action that were applied (set aside and crop rotation with non-irrigated crops). It is recommended in cases where actual data on water used for irrigation purposes are missing. The inclusion of IACS data at different points in time and linkage of individual land parcels to the farm will enable you to estimate the net effect exploring changes between different comparison groups (early-late joiners, drop outs etc.).

5.3.4 Indicator: Gross Nitrogen Balance

1. <u>Definition / description of the indicator, including:</u>

Environmental public good: Water quality
Type of indicator: Gross Nitrogen Balance is an impact indicator proposed by the CMES that indicates potential surplus of nitrogen on agricultural land.
Reflected RDP priority and focus area:

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.

- Focus area 4B: Improving water management, including fertiliser and pesticide management.

Unit of measurement: kg of nitrogen per ha per year

Type of data required: Gross Nitrogen Balance is estimated by a biophysical model using the nitrate fertiliser usage per land parcel, a detailed soil map and type of crop. LPIS-GIS data for participants and non-participants at different points in time are needed. Since IACS/LPIS data at different points in time are available, the deadweight effects could also be estimated.

Scale and level of application: from land parcel to regional and national level.

2. Existing data sources

EU-level: Data for EU-27 could only be compiled for 2005-2008 (Eurostat, GNB)

MS and regional level (examples): Gross nitrogen balances are not comparable between countries due to differences in definitions, methodologies and data sources used by countries. Nitrogen surplus (kg N/ha) is available for Norway and Switzerland between 1990 and 2008 (Nitrogen outputs, (kg N per ha), 1990-2008,EU-27, CH and NO)

Fact sheets and information available from other sources:

- RDP 2007-2013, CMEF, Guidance note J – Impact Indicator Fiches, Improvement in water quality
- IRENA indicator fact sheet 18.1 Gross nitrogen balance

3. Context of the case study testing

Case study area: Karditsa regional department, Thessaly Plain, Greece

Policy context: AE action for the reduction of nitrate pollution caused by agriculture in Nitrate Vulnerable Zones (NVZ)

Used data: IACS geo-referenced data of 2011 for participants and non-participants including the ha of supported area and type of crop, a soil map of a specific site of the plain area of Karditsa

Evaluation approach tested: Land parcel level and the specific site of the NVZ of Karditsa were the units of analysis for the micro and macro level, respectively. Given that IACS georeferenced data for participants and non-participants were only available for 2011, 2 comparison groups were constructed, using a naive counterfactual approach comparing only participants in the nitrate pollution reduction scheme and non-participants.

Table 13 Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Using the GNB in combination with a detailed soil map enables the consideration of specific environmental characteristics, such as soil texture and relief, as well as significant crops.	Missing information on farm characteristics (e.g. use of agricultural area, type of farming system, participation in RDP or other policy measures) can restrict the application of the indicator in elaborate statistics-based evaluations.	The impact of the AE action is estimated within each soil class.

Appropriateness of indicator(s) to capture complexity of environmental relationships	The GNB is strongly related to the nitrogen pollution from agricultural sources.	Actual data on fertiliser inputs are missing.	The impact of the AE action takes into account the different farming practices applied.

4. <u>Recommended application</u>

The GNB indicates the potential surplus of nitrogen (N) on agricultural land (kg N/ha/year) and also provides trends on nitrogen inputs and outputs on agricultural land over time. The specific indicator was estimated by a biophysical model and analysed in relation to the applied different farming practices of the AE action (set aside and crop rotation with non-irrigated crops). It is recommended in cases where actual data on fertiliser inputs are missing. The inclusion of IACS data at different points in time and linkage of individual land parcels to the farm they belong to will enable you to estimate the net effect, exploring changes between different comparison groups (early and late joiners, drop outs etc.).

5.3.5 Indicator: GNB for the assessment of effects of advisory services

1. <u>Definition / description of the indicator, including:</u>

Environmental public good: Water quality

Type of indicator: CMES impact indicator

Reflected RDP priority and focus area:

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
- Focus area 4B: Improving water management, including fertiliser and pesticide management.

Unit of measurement: amount of nitrogen per ha utilized agricultural area (kg N/ha)

Type of data required:

- Gross nitrogen balances for participating and non-participating farms that are calculated equally and reliably (to ensure comparability)
- Farm structural data, particularly on land-use such as share of grassland and main crops, and livestock density
- Reliable information on nitrogen cycles of farms to calculate robust gross nutrient balances (including import and export of organic N and purchases of feed and seed)

Scale and level of application: nitrogen balances are calculated at farm level

2. <u>Existing data sources</u>

EU-level: Data for EU-27 could only be compiled for 2005-2008 (update is planned for July 2016). As methodologies (especially with regards to the coefficients) and data sources used in different countries vary substantially, the balances are not consistent across countries, which means that data cannot be compared between countries (http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_gross_nitrogen_balance). To improve the comparability, harmonisation of methods and data sources in the member states is essential. The current situation also results in different qualities of evaluations due to different methods of calculations. Minimum requirements for data quality and the analysis should be defined to achieve comparability of good quality nitrogen balances.

***MS and regional level* (examples)**: Lower Saxony, Germany: Nitrogen balances are used to analyse the effect of advisory services (in combination with AE measures) on water quality. Database includes nitrogen balances of participating farms and a reference group generated from the controls of the fertiliser ordinance. An additional data set of 160 model farms was established in target areas of the Water Framework Directive (WFD). Recently, a reference group with farms outside of the WFD target areas was created. A description of the indicator (in German) is included in a handbook of the managing authority (NLWKN, 2015a)

Fact sheets and information available from other sources: Water quality impact indicator fiche by EU-Commission:

- Period 2007 – 2013: http://ec.europa.eu/agriculture/rurdev/eval/guidance/note_j_en.pdf
- Period 2014 – 2020: http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetailDoc&id=6707&no=3

3. Context of the case study testing

Case study area: Lower Saxony, Germany

Policy context: extension services addressing the protection of water resources are very important. Advisory services of farmers are supported and an own measure in drinking water protection areas was established. The main measures of RD programmes are: 1) 323 Rural heritage (support of technical advice in drinking water protection areas) and 2) 114 Use by farmers and forest holders of advisory services (farm management: focus on nutrient management).

Used data:

- Data from 160 model farms that receive intensive advisory service due to their location in target areas of Water Framework Directive are available for the years 2006 to 2012. As a reference group, farms of the fertiliser ordinance controls without any AE measure or advisory service are used.
- The data set of the controls of the fertiliser ordinance contains farms participating in advisory services related to drinking water protection. Farms that are not participating in this measure or in AE measures are used as the reference group.

Evaluation approach tested: Nitrogen balances are compared between participants and non-participants of advisory services related to the improvement of water quality. The classic approach is used with two comparison groups for each analysis. Non-participants are compared with farms that receive advisory services. For the first test (combination of model farms with data of the controls of the fertiliser ordinance), two data sets that stem from different sources are combined. Thus, before conducting comparative statistical analysis, the two data sets have to be tested for structural differences. It turned out that the comparability of the two different data sets is limited. Recently, a reference group for the model farms was constructed and a comparative analysis was conducted by the monitoring organisation (NLWKN). Results show reduction effects of advisory services on the nutrient balances of farms with advisory service over the years while the nutrient balances of non-beneficiaries remained static (NLWKN, 2015b). However, this data was not available at the time of the analysis. Thus, a comparison of participants with non-participants using the control data set of the fertiliser ordinance was conducted. Propensity Score Matching is used to match similar farms. As gross nutrient balances are not included in the data set of the fertiliser ordinance controls, net nutrient balances are used for the comparative analysis.

Table 14 1. Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Typology according to the available data on environmental conditions and farm structure is used - Grassland share - Share of arable land - Amount of organic N at farm level Matching of farms with similar characteristics (N removal, N application and amount of organic N at farm level) improves the robustness of results	Good quality data for reference group is rarely available, e.g. here only net nutrient balances are available for the reference situation and data is only available for one point in time (no panel data) Limited information on farm structural data and management practices, e.g. information on livestock density and type of main crops is missing for participants Information on intensity of advisory services needed to further improve the assessment Nitrogen balances are based on calculations	Testing of use of data from different data sources to construct robust counterfactual Use of a matching approach to compare similar farms
Appropriateness of indicator(s) to capture complexity of environmental relationships	Indicator has a strong linkage to the potential nitrate pollution of the groundwater Indicator is well known and used in the case study area	Indicator is a proxy for groundwater quality. It takes long time until effects can be measured in the groundwater Assessment of net-effects of advisory service is difficult due to joint implementation with AE measures. Comparability of data from different sources is limited	Quantitative assessment of water quality impacts of advisory measures

4. <u>Recommended application</u>

GNB indicator provides reliable information on nitrogen management of farms and is widely used for the analysis of nutrient surpluses and impact assessment of AE measures due to its explanatory power (NLWKN, 2015a)

GNB indicator can also be used to assess the effects of advisory services related to water quality using a similar approach as for the analysis of AE measures

Net-nutrient balances are based on rough estimations on the amount of organic N as well as estimations of forage and grassland yields. Therefore, gross nutrient balances are more reliable and should be favoured over net balances.

To enable the application of advanced statistics based assessments of net-effects data should include sufficient information on gross nutrient balances and farm structure and management practices for participating and non-participating farms. Minimum requirements are:

- Gross nutrient balances should be available for both groups

- Reliable information on the components of the Gross Nitrogen Balance which cover the whole nitrogen cycle of the farm
- Information on main crops and grassland share as well as livestock density
- Information on type of AE and other RD measure
- Panel data for participants and non-participants should be available

In addition to the nitrogen balances, single components of the balance, e.g. the amount of mineral fertiliser purchased, can also be used for impact assessment (NLWKN, 2015a)

5.4 Landscape

5.4.1 Method: Spatial analysis

1. Definition / description of the method, including:

Type of method: Spatial analysis

Environmental public good: Landscape

Micro or macro level application: The availability of representative data at the farm level would allow for aggregation at the macro level.

2. General requirements

Data requirements: UAA (IACS, FADN, etc.), land cover/land use data, remote sensing and aerial photography data, landscape and vegetation maps

Skill requirements: Spatial analytical /GIS skills.

3. Consideration of counterfactuals

The method can be used for Qualitative and Naive Quantitative Evaluation Options and Statistics-based Evaluation Options. The quality and quantity of data play the key role in the construction of comparison groups. Thus the data availability on participants and non-participants before and after measure implementation determines what counterfactual approach will be used.

4. Context of the case study testing

Case study area: Northern Santorini, Greece

Policy context: 1.) Agri-environmental measure for the landscape protection: action for the maintenance of pruning and propagation practices in vineyards on the island of Santorini, 2.) Special measures in favour of the Small Aegean Islands concerning certain agricultural products, such as the continued cultivation of traditional vines.

Used data: Land-cover data of Northern Santorini drawn from GE images (2003 and 2012), IACS geo-referenced data of 2011 for participants.

Evaluation approach tested: Given that land-cover data were available between two periods in time (2003 and 2012), while IACS data provided only data for participants in 2011, three comparison groups were constructed, participants in the AE action, participants in the AE action plus the special aid to the Aegean islands, and participants in neither scheme.. A conceptual DiD approach was selected comparing the changes in traditional vineyards among the three comparison groups from 2003 to 2012.

Table 15 Strengths and weaknesses of the method (based on SWOT)

Evaluation challenges (relevant for indicators)	Strengths	Weaknesses	Key contribution to evaluation benefits
Compatibility with local environmental and farm structural characteristics	Local environmental characteristics are drawn from Google Earth imagery. GE provides free and user-friendly access to satellite images of varying resolution of the Earth's surface.	Manual digitisation and interpretation are time consuming and intuitive processes. Specific landscape features such as terraces and boundary walls are not visualised.	Changes in landscape are easily distinguished through the images.
Timing of environmental impacts captured	Image interpretation is based on elements that are inherent in GE imagery. Thus changes over time are very easily observed and captured.	The resolution of historical images may be too coarse for detailed mapping.	IACS georeferenced data are theoretically available every year. GE images vary by area and time.
Establishment of robust causal relationships	Method provides quantitative information but is not able to explain the effects.	Land-cover maps produced were not tested in their entirety for accuracy. Neither statistical tests nor regression analysis were conducted	Method is based on well-documented, theoretically sound models that could link farming practices with the environmental outcomes.
Establishment of consistent micro-macro linkages	IACS georeferenced data and land-cover data are spatially explicit; thus the aggregation of all land parcels forms the landscape of Santorini. Changes are visualised at land-parcel level as well as at landscape level.	Farm level, which is the decision level for participation in the various schemes, was missing. The functional unit was not linked to a programme scale.	Macro-level analysis can be built on aggregated micro-level results. This approach establishes linkages with biophysical variables that are suitable to upscaling.
Assessment of net-impacts	The estimation of direct and indirect effects is based on the availability of control groups.	Partial information on participants and non-participants before and after the implementation of the AE action is available.	DiD analysis is limited only to the changes observed.
Appropriateness of indicator(s) to capture complexity of environmental relationships	Method can capture the complexity of the environmental relationships and is used for monitoring land use and land cover changes.	Land-cover maps produced were not tested in their entirety for accuracy. A lot of interpretation errors were identified during the ground-truth survey.	Method relies on spatial data on land cover. Ground-truth survey data can be used to address data gaps.
Unambiguous and understandable results	Results are easy to communicate to laypersons.		Method provides user friendly outcomes.

6. Recommended application

Given that the techniques of photo-interpretation and manual digitisation are time consuming, the method selected seems to be suitable for small and site-specific schemes, which are applied in limited and defined areas with unique landscape characteristics Since the method uses spatially-explicit data, causality between rural development interventions and changes in landscape is improved. In terms of data requirements, the IACS-LPIS data set on participants and non-participants covering a long time series and also linkages between land parcels and the farm are necessary in order to use elaborated statistics counterfactual explaining in more depth the link between policy interventions and changes observed.

5.4.2 Method: Landscape metrics

1. **Definition / description of the method, including:**

Type of method: Spatial analysis

Environmental public good: Landscape

Micro or macro level application: Landscape Metrics (Botequilha Leitao, et al 2006) are based on landscape ecological principles (Farina, 2007), which include indicators along a common scale that measures patterns/structures in a landscape, from patch (an area of single land use/land cover), through class (total area of single land use/land cover) to landscape level. The method, therefore, has embedded the micro and macro levels. However each level has its own a range of indicators. The same methodological process will generate both micro and macro-level results.

2. **General requirements**

Data requirements: agricultural land use data (IACS, LPIS), land cover data (CORINE land cover or national equivalents). The method is sensitive to data scale; hence data in an assessment need to be of compatible resolutions for a comparison of results.

Skill requirements: spatial and statistical analytical skills, sound GIS skills.

3. **Consideration of counterfactuals**

Landscape metrics measure a landscape as a whole with micro and macro-level indicators. The creation of comparison groups requires the creation of individual land use/cover data layers for the case study area for each part of the comparison group (with and without). The potential for use of an elaborate statistical approach requires acceptable explanatory factors incorporated in the comparison groups. In the absence of explanatory factors, a naïve estimate of counterfactual (Difference-in-Difference) can be used.

4. **Context of the case study testing**

Case study area: Grampian region, Scotland

Policy context: During the RDP 2007 – 2013 AE actions aimed to safeguard and enhance landscape; native woodland; non-native woodland; and geo-diversity. The main measures were: agri-environment (214), woodland creation (223) and woodland management (225).

Used data: IACS geo-referenced land-use data for 2008-2014 for participants in measures 214, 223 and 225 and non-participants, Land Cover Map 2007 (CEH, 2011) to fill data gaps, and the National Forest Inventory (Forestry Commission).

Evaluation approach tested: Land-use patches and the case study area as a whole were the unit of analysis for the micro and macro level, respectively. The data allowed for the creation of comparison groups of before and after as well as with and without participation. However the data were insufficient to explain the participation for an elaborate statistics-based evaluation; instead a naïve Difference-in-Difference counterfactual approach was used.

Table 16 Strengths and weaknesses of the method (based on SWOT)

Evaluation challenges	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Local characteristics are represented by the land use or land-cover data which form the basis of the assessment	The method is sensitive to scale, resolution and quality of input data.	Landscape metrics provide ways of assessing impacts of RDP on landscape and HNV based on unique indicators at different levels.
Timing of environmental impacts captured	The use of IACS land use data is updated annually for every field or land parcel.	The change that has an impact on landscape introduced by RDP measures is often sub-field. In current analysis impacts are under estimated.	Landscape metrics can be applied on data on land use or cover captured to best assess the rate of change in environmental impacts, with and without RDP measures.
Establishment of robust causal relationships	The method is based on a theoretically sound basis of measuring and monitoring landscape change using land cover/use data.	IACS land use data used were the best available at for individual land parcels, however more spatially refined land cover data would support the establishment of more robust causal relationships.	Landscape metrics support the derivation of robust causal relationships at different scales, directly related to RDP measures, to assess net impacts at macro-levels.
Establishment of consistent micro-macro linkages	The unique indicators for micro (patch) and macro level (landscape) in this method are linked by their explicit relationship: individual patches of land use make a class, and different classes comprise a landscape.	The method has in-built consistency but the indicator is not a single metric.	The method is based on consistent micro-macro linkages.
Assessment of net-impacts	The estimation of direct and indirect effects requires the availability of control groups.	Counterfactual approach limited to naïve DiD by data constraints. Explanatory data for different comparison groups were missing and need more testing.	Supporting the assessment of net-impacts at macro level.
Appropriateness of method to capture complexity of environmental relationships	Landscape metrics assess at different levels the structure/patterns in the landscape	The method is scale dependent, hence sensitive to data quality.	The landscape metrics method provides results which are relevant to site specific environmental conditions.
Unambiguous & understandable results	Results can be communicated to laypersons.		Method provides user friendly outcomes.

5. Recommended application

The landscape metric method is recommended for the assessment of macro level impacts of RDP on changes in landscapes and HNV. The method introduces an impact assessment of RDP driven land use change in the context of its surroundings areas which is important, particularly for public goods biodiversity (HNV) and landscape. The quality of land cover data may be a constraint on an impact assessment, however current developments in remote sensing (e.g. Copernicus Programme) and hand held technology could address some of these limitations.

5.4.3 Indicator: Shannon Diversity Index

1. Definition / description of the indicator, including:

Environmental public good: Landscape

Type of indicator: Shannon Diversity Index (SDI) is a proposed additional impact indicator. The SDI is most commonly used for the assessment of ecological diversity; however it is also applied for the assessment of landscape diversity.

Reflected RDP priority and focus area:

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
- Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.

Unit of measurement: proportion of landscape occupied by patch type

Type of data required: land use/land cover data, IACS/LPIS

Scale and level of application: local, regional and national areas.

2. Existing data sources

EU-level: Data for EU-27 is available through the CORINE land cover data for 2006 only (http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3).

***MS and regional level* (examples)**: at national level land cover data may have local modifications but generally these can be reclassified into CORINE classes to facilitate comparison between MS. For the UK this is Land Cover Map 2007 (LCM2007).

Fact sheets and information available from other sources: none, but linked to the EU landscape convention.

3. Context of the case study testing

Case study area: Aberdeenshire, Scotland

Policy context: Agri-environmental and forest-environment measures include objectives for enhancements of the rural landscape which will have a positive impact on people's environment and highlights the importance of Scotland's woods and forests.

Used data: IACS/LPIS geo-referenced land use data of 2008-2014 for participants and non-participants, and LCM2007 to fill the gaps in the IACS land use data.

Evaluation approach tested: Individual areas of single land use/land cover (i.e. patches) were created for baseline data (before) and following years under RPD. Comparison groups were created by identifying change in LULC against the baseline for areas with and without RDP participation. Shannon Diversity Index was calculated for the two separate comparison groups and compared against the baseline. While the data allowed for an assessment of before and after as well as with and without participation, the data were insufficient for an elaborate statistics-based evaluation. Instead, a naive counterfactual approach was used based on simple means.

Table 17 Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Local environmental characteristics are used through land use or land cover data which form the basis for the assessment	The method is scale sensitive, which means that the data quality determines	Landscape metrics as a method introduces ways of assess impact of RDP on landscape and HNV based on unique indicators at different levels.
Appropriateness of indicator(s) to capture complexity of environmental relationships	Shannon Diversity Index is an established indicator to capture complexity	The indicator is scale dependent, hence sensitive to data quality	Shannon diversity index as part of the landscape metrics method provides results in relation to the site specific environmental conditions.

4. Recommended application

The Shannon Diversity Index is a common way to assess the structural complexity/diversity of an area at macro level (landscape) and is able to provide trends on the characteristics of the structure of agricultural landscapes over time. The quality (data resolution) of the data used to calculate the Shannon Diversity Index for different comparison groups needs to be consistent because the indicator is scale sensitive. Among the landscape metrics there are a range of different diversity indicators; however the Shannon Diversity Index is the best known and most commonly used.

5.4.4 Indicator: Patch shape index

1. **Definition / description of the indicator, including**:

Environmental public good: Landscape
Type of indicator: patch shape index is a proposed additional programme specific result indicator, which measures the geometric complexity of a patch (i.e. an area of the same land use/cover) and the impact of RDP on the landscape.
Reflected RDP priority and focus area:

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
- Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.

Unit of measurement: ratio of patch perimeter (m) and the square root of patch area (m^2), adjusted by a constant to adjust for a square standard.

Type of data required: land use/land cover data, IACS/LPIS
Scale and level of application: patch and landscape level

2. Existing data sources

EU-level*: Data for EU-27 is available through the CORINE land cover data for 2006 only (http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3).*
MS and regional level (examples)*:* at national level land cover data may have local modifications (for case study, LCM2007) but generally these can be reclassified into CORINE classes to facilitate comparison between MS
Fact sheets and information available from other sources: none, but linked to the EU landscape convention.

3. Context of the case study testing

Case study area: Aberdeenshire, Scotland
Policy context: Agri-environmental and forest-environment measures include objectives for enhancements of the rural landscape which will have a positive impact on people's environment and highlights the importance of Scotland's woods and forests.
Data used: IACS/LPIS geo-referenced land use data of 2008-2014 for participants and non-participants and LCM2007 to fill the gaps in the IACS land use data.
Evaluation approach tested: Individual areas of single land use/land cover (i.e. patches) were created for baseline data (before) and following years under RPD. Comparison groups were created by identifying change in LULC against the baseline for areas with and without RDP participation. Patch Shape Index calculated for the two separate comparison groups against the baseline could be compared. While the data allowed for an assessment of before and after as well as with and without participation, the data were insufficient for an elaborate statistics based evaluation, instead a naive counterfactual approach was used based on simple means.

Table 18 Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Local environmental characteristics are used through land use or land cover data which form the basis for the assessment	The method is scale sensitive, which means that the data quality determines	Landscape metrics as a method introduces ways of assess impact of RDP on landscape and HNV based on unique indicators at different levels.
Appropriateness of indicator(s) to capture complexity of environmental relationships	Patch shape index is one of the indicator able to capture the structure/patterns of the landscape	The indicator is scale dependent, hence sensitive to data quality	Patch shape index as part of the landscape metrics method provides results in relation to the site specific environmental conditions.

4. Recommended application

The Patch Shape Index (PSI) is one of the ways to assess the structural complexity of an area (at micro i.e. patch level). It is able to provide trends on the characteristics of the structure of agricultural landscapes over time by comparing change to a baseline. The change can be attributed to participants and non-participants to RDP measures. PSI calculated for separate comparison groups against the baseline can be compared. The data quality (data resolution) of the data used to calculate the PSI for different comparison groups need to be consistent because the indicator is scale sensitive.

5.4.5 Indicator: Land-cover change

1. Definition / description of the indicator, including:

Environmental public good: Landscape

Type of indicator: Additional programme specific indicator. Moreover, land-cover change is a pressure indicator based on IRENA operation (No24) that identifies land-cover changes to and from forest/semi-natural and agricultural land (EEA Report, 2006).

Reflected RDP priority and focus area:

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
- Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.

Unit of measurement: Changes in land cover classified by type and size (%)

Type of data required:

- LPIS-GIS and land-cover data between participants and non-participants before and after measure implementation at land parcel level.
- Also, including information on farm structural variables (e.g. use of agricultural area, yields, type of farming system, input cost, participation in RDP or other policy measures) will enable you to assess the indirect effects, such as the deadweight effects, by checking change/maintenance observed in non-participants that would have occurred even in the absence of the applied measures.

Scale and level of application: spatial and temporal scale, NUTS 2/3 level (where data is available).

2. Existing data sources

EU-level: Since 2006, EUROSTAT has carried out a survey on the state and the dynamics of changes in land use and land cover in the European Union; this is called the LUCAS survey (Land Use/Cover Area Frame Statistical Survey). LUCAS is a field survey based on an area-frame sampling scheme. Data on land cover and land use are collected, and landscape photographs are taken to detect any changes to land cover/use to European landscapes. These surveys are done every three years.

***MS and regional level* (examples)**: There are two main types of information derived from LUCAS: aggregated statistical data and elementary data (for individual survey points). The aggregated results show land cover and land use for the EU-27 and national averages for the EU Member States, and can also be shown at a more detailed level, for example, for more than 250 NUTS 2 regions. Moreover, relevant work published by WWF-Greece presents the spatial data on land cover and its change tendencies in Greece from 1987 to 2007. (http://www.wwf.gr/en/areas/forests/land-uses).

3. Context of the case study testing

Case study area: Northern Santorini, Greece

Policy context: 1. AE action for the maintenance of pruning and propagation practices in vineyards on the island of Santorini, 2. Special measures in favour of the Small Aegean Islands.

Used data: Land-cover data of Northern Santorini drawn from GE images (2003 and 2012), IACS geo-referenced data of 2011 for participants in the AE action and the special measure.

Evaluation approach tested: Land parcel level and landscape of Northern Santorini were the units of analysis for the micro and macro level respectively. Land-cover data were available between two periods in time (2003 and 2012), while IACS data provided only for participants of 2011. Due to data availability, three comparison groups were constructed, participants in the AE action, participants in the AE action and the special aid to the Aegean islands and participants in neither scheme. A conceptual DiD approach is selected comparing the changes in traditional vineyards among the three comparison groups from 2003 to 2012.

Table 19 Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	Strengths	Weaknesses	Key contribution to evaluation benefits
Compatibility with local environmental and farm structural characteristics	Land-cover data are drawn from GE satellite images from 2 different years (2003 and 2012) paying particular attention to the area covered with vines pruned using traditional pruning techniques. These vineyards are the key features that form the unique landscape of Santorini.	The classified land-cover polygons consist of more than one land parcel, since the manual digitisation was processed taking into account neighbouring features (i.e. the adjacent land parcels with the same spatial characteristics were grouped to one polygon). Thus in some cases it is difficult to estimate the precise number of land parcels per classified polygon.	Changes in land cover and in particular changes in traditional vineyards are suitable for monitoring the implementation of the AE measure.
Appropriateness of indicator(s) to capture complexity of environmental relationships	Indicator is used for monitoring land use and land -over changes.	The land-cover maps produced have not been tested for accuracy in their entirety. Many interpretation errors were identified, during the ground truth survey conducted.	Indicator can provide useful information on changes in traditional vineyards over time.

4. Recommended application

The indicator selected seems to be suitable for very site-specific schemes which are applied in limited and defined areas with unique landscape characteristics. Changes in vineyards were easily distinguishable due to their spatial characteristics, thus the proposed indicator is appropriate to permanent, not extended crops with unique characteristics. The inclusion of IACS georeferenced data for participants and non-participants before and after measure implementation is considered the minimum required data for estimating this indicator. In cases where information except for the land parcel level is also connected to the farm, it is possible to estimate the net impact.

5.4.6 Indicator: Visibility of change

1. Definition / description of the indicator, including:

Environmental public good: Landscape

Type of indicator: Visibility of change is a proposed alternative impact indicator which captures the changes in visibility of individual patches due to uptake of RDP measures.

Reflected RDP priority and focus area:

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
- Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.

Unit of measurement: Area of visible land-use change; proportion of visible land-use change within landscape character area

Type of data required: digital terrain model, IACS/LPIS data, and topographic data,

Scale and level of application: From feature to regional, landscape character and national level

2. Existing data sources

EU-level: Data of relevant scale and spatial resolutions are not held at an EU level

***Member State and regional level* (examples)**: Relevant Member State data: IACS RDP uptake, spatial units (i.e. generally field boundaries), national topographic mapping (Ordnance Survey MastermapTM), Digital Terrain Model (1:10,000) and Landscape Character Assessment mapping.

Fact sheets and information available from other sources: No fact sheets are available, but further information on the indicator can be obtained from other scientific publications listed under references below.

3. Context of the case study testing

Case study area: Grampian region, Scotland.

Policy context: Agri-environmental and forest-environment measures aimed to safeguard and enhance the landscape and its character; native woodland and associated habitats and species; non-native woodland and associated habitats and species; and geo-diversity.

Used data: IACS geo-referenced data of 2009 and 2014 for participants and non-participants including the type of crop, National Forest Inventory, topographic data (Ordnance Survey MastermapTM) and Digital Elevation Model (DEM) data, Landscape Character Assessment.

Evaluation approach tested: The land parcel is the unit of analysis for the macro-level analysis. At Step 1, for the case study area, a baseline of the visibility of land-cover types is calculated to enable comparisons of before and after. Information on the types of features associated with the uptake of RDP measures is used as input for the land parcel. At Step 2, the macro level analysis is carried out with respect to the visibility of individual units, which are cumulated to landscape character areas. At Step 4, the outputs can be presented as time series, and different groupings of land parcels selected to show impacts with and without uptake, stratified according to size, distribution (e.g. clustered or distributed according to a specified pattern geographically, or temporally), and interpreted with respect to landscape character map units.

Table 20 Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Generalised local environmental characteristics are used, derived from other data on land cover and use, but with compatible classifications and geographic scale	Information on the surrounding vegetation which provides a visual context (e.g. colour, texture, shape) and so contracts with the vegetation change due to RDP measures are not easily quantified and thus used in the calculation. Therefore, visibility is assumed due to presence of the patch/feature in the view irrespective of the contrast with the background vegetation, weather and other ephemeral conditions.	The indicator uses inputs which are directly related to RDP measures and uptake, with the change in landscape related to the context of the characteristics of the surrounding landscape.
Appropriateness of indicator(s) to capture complexity of environmental relationships	Direct link to widely-used definition of landscape areas (LCA), with easily understood interpretation (visibility of features), using an indicator which represents a clear impact on, or contribution to the public good of landscape, with a theoretical basis which provides causal links. Repeatable method.	Interpretation required with respect to landscape character to assess the net effects on landscape, thus requiring qualitative judgement by expert or following relevant training. Time due principally to intensive computer processing requirements.	The indicator provides a direct measure of the impact on the visual landscape of RDP measures. The interpretation with respect to landscape character and through principles of theory provides an understanding of the causal links between the type and extent of change and the impact on landscapes.

4. Recommended application

The Visibility of change indicator identifies the impact of RDP-driven land-use change on the landscape and its character. Calculations of the change in visibility can be assessed annually, and compared to a baseline pre-uptake, so enabling the identification of trends through time.

5.4.7 Indicator: Visual amenity

1. <u>Definition / description of the indicator, including:</u>

Environmental public good: Landscape

Type of indicator: Additional programme-specific indicator adapted by the research team for measuring the amenity values offered by the traditional vineyards. Given that there is a causal link established between Santorini's traditional vineyards and its contribution to making this landscape attractive for people, we assume that the alteration of iconic vineyards resulting from the characteristic pruning systems will deteriorate its distinctive landscape and consequently people's amenity will be affected negatively. The estimation of visual amenity indicator is categorised into three levels of an ordinal scale, high-medium-low. High level is considered when the pruning system is maintained, medium when land cover (vineyard) is maintained, low when the land cover is changed. The allocation of the visual amenity level to each comparison group is based on the observed changes in area of vineyard in the timeframe 2003-2012.

Reflected RDP priority and focus area:

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
- Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.

Unit of measurement: loss of amenity values based on the observed land cover changes from 2003 to 2012 (%)

Data requirements: The specific indicator is based on land cover changes observed in area of traditional vineyards in the timeframe 2003-2012. Thus LPIS-GIS and land-cover data between participants and non-participants at different points in time are required.

Scale and level of application*:* spatial and temporal scale, from land parcel to regional level

2. <u>Context of the case study testing</u>

Case study area: Northern Santorini, Greece

Policy context: 1. AE action for the maintenance of pruning and propagation practices in vineyards on the island of Santorini, 2. Special measures in favour of the Small Aegean Islands.

Used data: Land cover data of Northern Santorini drawn from GE images (2003 and 2012), IACS geo-referenced data of 2011 for participants.

Evaluation approach tested: Land parcel level and landscape of Northern Santorini were the units of analysis for the micro and macro level respectively. Given that land-cover data were available between two periods in time (2003 and 2012) and IACS data provided only for participants of 2011, three comparison groups were constructed, participants in the AE action, participants in the AE action and the special aid to the Aegean islands and participants in neither scheme. A conceptual DiD approach was selected comparing the amenity values based on their observed land cover changes from 2003 to 2012 among the three comparison groups.

Table 21 Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Vineyards pruned with the traditional techniques are the key features that form the unique landscape of Santorini producing high scenic beauty.	The categorisation of indicator into three levels is based on arbitrary criteria.	The main objective of the AE action explicitly state that vineyards pruned with the traditional techniques offer high visual quality in Santorini's landscape.
Appropriateness of indicator(s) to capture complexity of environmental relationships	In the case of Santorini, visual amenity is based on an intuitive interpretation of how land-cover changes may affect the amenity values offered by traditional vineyards.	Measurement of indicator is based on a subjective method.	The main objective of the AE action explicitly state that vineyards pruned with the traditional techniques offer high visual quality in Santorini's landscape.

3. Recommended application

Given that many scientific studies have been performed on the visual quality of landscapes and how the agricultural landscape could be evaluated, the research team built a visual quality indicator. In the case of Santorini's traditional vineyards, the AE action explicitly states which landscape offers high amenity values, i.e. a more natural distribution of the traditionally pruned vines. Assuming that changes in land cover have an impact on the attractiveness of Santorini's vineyards, visual amenity analysis was based on intuitive interpretation of land-cover changes observed in traditional vineyards from 2003 to 2012. Thus land-cover data and IACS/LPIS data between participants and non-participants before and after measure implementation are required. However the indicator selected is based on the arbitrary assignment of values to land-cover types. The risk of non-comparability holds for the specific selection.

5.5 Biodiversity HNV

5.5.1 Indicator: High Nature Value forestry

1. Definition / description of the indicator, including:

Environmental public good: Biodiversity High Nature Value (HNV) forestry

Type of indicator*: The indicator shows the increase or decrease of ecotone length between afforested and adjacent land.*

Reflected RDP priority and focus area*:*

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
- Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.

Type of data required: IACS georeferenced data, Forest cadastre data, Orthophoto image, land-use data.

Scale and level of application: Each land parcel of each relevant RDP measure at micro level and the group of land parcels of each relevant RDP measure in the geographical region at macro level.

2. Existing data sources:

EU-level: IACS data, orthophoto image, land-use data

***MS and regional level* (examples)**: IACS data is available through national rural agencies.

Fact sheets and information available from other sources: -

3. Context of the case study testing

Case study area: Nevezis wooded agrarian and urbanised plain area, Lithuania

Policy context: RDP measures: 214 Agri-environment payments, 221 First afforestation of agricultural lands, 223 First afforestation of non-agricultural lands. Since forest land is rapidly increasing on abandoned land, the ecotone[10] would be a good indicator to measure the afforestation measure implementation in the area.

Used data: IACS-LPIS georeferenced data, Forest cadastre database, Georeferenced spatial data set at 1:10,000 scale of the Republic of Lithuania (GDR10LT), Orthophoto image 2010-2014.

Evaluation approach tested: Using the before-after approach, the changes in the length of the ecotone was calculated in the selected area, to see what impact the application of RDP measures had on the heterogeneity of landscapes. At micro level, randomly-selected land parcels (as accounted in LPIS databases) were chosen for the calculations to see if the applied measure has a positive or negative effect on the heterogeneity of landscapes. The problem with the application of the method at micro level was that it is not an automated process and it is time intensive. Every parcel had to be calculated and reviewed manually. At macro level, the method does not calculate the effect of each land parcel but the effectiveness of the measure as it showed the consolidated

[10] An ecotone is a transition area between two biomes. It is where two communities meet and integrate. It may be narrow or wide, and it may be local (the zone between a arable field and forest) or regional (the transition between forest and grassland ecosystems). An ecotone may appear on the ground as a gradual blending of the two communities across a broad area, or it may manifest itself as a sharp boundary line (Fagan et al. 2003, Hansen et al. 1988; Wiens 1992).

results from all the parcels in the region. The effectiveness was measured by extension or decrease of the ecotone length. This process was an automated one. The main problem was that the forest cadastre data is only renewed once every 10 years and it sometimes produces discrepancies with the IACS data. Also it should be mentioned that the results from this indicator should not be considered in isolation before concluding whether the measure had a positive or a negative effect. Other parameters, like habitat connectivity, habitat patching, should also be considered and the evaluation should only be concluded from consolidated results.

Table 22 Strengths and weaknesses of the indicator

Evaluation challenges (relevant for indicators)	Strengths	Weaknesses	Key contribution to evaluation benefits
Compatibility with local environmental and farm structural characteristics	Easy to make comparison, because the new ecotones are easy to monitor not depending on the scale	For quality impact (this indicator only shows the changes of ecotone in length, the quality is not assessed) assessment, there is a need to collect data on site because available datasets are inadequate for such assessment	Spatial changes in landscape and HNV territory are easily distinguished
Timing of environmental impacts captured	The spatial change of the ecotone is captured after the measure is implemented	For quality assessment at least 5 years permanent monitoring is needed to collect additional data	Spatial changes can be captured every year because IACS-LPIS georeferenced data are available every year
Establishment of robust causal relationships	The calculation of the indicator shows direct influence in the quantity of ecotone	To be able to assess impact on quality additional research and data are needed	The application of measures show direct influence on landscape heterogeneity changes
lishment of consistent -macro linkages	Methodological approach explicitly covers and combines micro and macro level analysis. Consistency and validation procedures are internalised.	None	Macro-level analysis can be built on aggregated micro-level results. This approach establishes linkages with variables that are suitable to upscaling.
Assessment of net-impacts	None	This method is not a stand-alone method. To assess net-impacts, other indicators have to be taken into account.	Before and after analysis is limited only to the changes observed.
Appropriateness of indicator(s) to capture complexity of environmental relationships	It is easy to capture one aspect – RDP impact on heterogeneity of landscapes	It is not enough to use the proposed indicator to be able to measure complexity of environmental relationships	Method relies on spatial data on application of measures. Ground truth survey data can be used to address data gaps.
Unambiguous and understandable results	Results are easy to understand and communicate, no specific technical skills are required	To present complexity of environmental relationships more results from other indicators are needed as this indicator as single is not sufficient	Method provides user friendly outcomes.

4. **Recommended application**

This method is applicable for very site-specific schemes. It shows the best and most effective results at micro level, as there you can count what the precise effect of the RDP measure will be if it is applied in the area or not. However, it also helps in providing information on the overall situation at the macro level. For better usage of the method at micro level, the problem of automatisation needs to be solved, as well as the timing of updating the different databases. The minimum requirements to use this method are not large. The person should have basic GIS skills.

5.5.2 Indicator: High Nature Value farmland

1. **Definition / description of the indicator, including:**

Environmental public good: Biodiversity HNV farmland

Type of indicator: High Nature Value (HNV) farmland is typically characterised by a combination of low intensity land use, the presence of semi-natural and unfarmed features and a diversity of land cover and land uses, supporting the presence of high-level biodiversity of wildlife species and habitats. HNV farmland and HNV farming systems are composite indicators. The basic components of these indicators are represented by: 1) high proportion of semi-natural vegetation; 2) mosaic of low-intensity agriculture; 3) supporting wild species and habitat of conservation concern.

Reflected RDP priority and focus area:

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
- Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.

Unit of measurement: a) Percentage of HNV farmland on UAA and b) HNV score at farm level

Type of data required: In case of measurement a): georeferenced data on land cover and land use with sufficient details to guarantee the assessment of semi-natural features, the level of farming intensity and the presence of wildlife species and the possibility of comparison between participants and non-participants (e.g. LPIS database). In case of measurement b): individual data of samples of farms with information on crops, livestock and type of farming practices.

Scale and level of application: HNV farmland may exist at different scales from single parcel to an entire landscape, while HNV farming system refers to land cover and associated farming practices of the system as a whole, either it is at farm level or at landscape level.

2. **Existing data sources**

EU-level: FADN, IACS, LPIS, JRC maps on HNV and semi-natural vegetation

***MS and regional level* (examples)**: Data on land cover, farming intensity (nitrogen and pesticide), and ecological quality index are available through the Regional Environment Agency (ARPAV) and Managing Authority (Veneto Region). Farmland Bird Index data from National Rural Network

Fact sheets and information available from other sources: Paracchini et. al, 2009; ENRD 2010; Keenleyside et al. 2014.

3. <u>Context of the case study testing</u>

Case study area: Veneto Region - Italy.

Policy context: Natural handicap payments to farmers in mountain areas (211) Agri-environment measures aimed to increase biodiversity 214/A (Ecological corridors, buffer strips, hedgerows and thickets), 214/C (Organic farming); 214/D (Protection semi-natural habitats and biodiversity), 214/E (Meadows and grasslands); 214/F (Biodiversity) and Support for non-productive investments (216)

Used data: IACS, LPIS, FADN, Land cover map, Farming intensity (nitrogen and pesticide) and Farmland Bird Index data

Evaluation approach tested: The quantification of HNV farmland and the assessment of the contribution of RDP measures to improve the diffusion of HNV farmland has been tested with the indicators (Percentage of Utilised Agricultural Area farmed to generate High Nature Value and Farms with high percentage (score) of HNV farmland) calculated in two steps: 1) identification of HNV farmland and 2) evaluating the capacity of the RDP to preserve and enhance HNV farmland. Multicriteria analysis has been extensively used to create composite indicators that summarise many different aspects of HNV farmed land measured with specific unit of measurement, and aggregated with the normalisation procedure. At micro level, an elaborate statistics evaluation approach can be applied if the sample of farms has a reasonable representativeness of participants and non-participants. At macro level, spatial analysis concerning participants and non-participants is applicable if the IACS-LPIS databases are available at cadastral level.

Table 23 Strengths and weaknesses of the indicator

Evaluation challenges (relevant for methods)	Strengths	Weaknesses	Key contribution to evaluation benefits
Compatibility with local environmental and farm structural characteristics	The score approach allows for the use of selected sub-indicators potentially specific and reflecting local environmental and farming conditions.	A better data set about landscape features and hedges distributed at farm and landscape level would be advisable.	The two-tier approach can investigate the differences of local contexts at micro level along with an overall picture at macro level.
Appropriateness of indicator(s) to capture complexity of environmental relationships	Composite indicators based on few or several sub-indicators can better assess the multiple definitions of HNV farmland	Difficulties to create comparable statistics among regions or Member States	The two-tier approach can investigate the differences of local contexts at micro level along with an overall picture at macro level.

4. Recommended application

The availability of a farm sample updated annually, such as FADN, gives the chance to monitor over time the evolution of HNV farmland at micro level. The representativeness of the FADN sample should be available at territorial level in order to ensure a greater consistency between micro and macro level. This could increase the number of observations needed to have a sufficient statistical significance of the estimated parameters required for an assessment of net-effects and, consequently, the cost of the analysis. The poor availability of data on the extent of semi-natural features in the farms could undermine the measurement of biodiversity values of a farmed area. The increasing availability of data concerning large and small patches of perennial vegetation detected in fine-resolution satellite images should increase the reliability of land cover in agro-ecosystems at reasonable monitoring costs.

5.6 Biodiversity Wildlife

5.6.1 Indicator: Number of farmland bird individuals

1. <u>Definition / description of the indicator, including:</u>

Environmental public good: Biodiversity Wildlife

Type of indicator: Additional programme-specific indicator

Reflected RDP priority and focus area:

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
- Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.

Unit of measurement: Number of farmland bird individuals

Type of data required: Regularly collected biodiversity related (bird census) data from previously set geological location in a timescale of the programme period under consideration

Scale and level of application: Biodiversity related data is evaluated at the level of the survey points of the observation

2. <u>Existing data sources</u>

EU-level: Data for the number of farmland bird individuals is the baseline data for Farmland Bird Index. Data collection (monitoring) standards of the common bird species is set by the European Bird Census Council. Results of the Farmland Bird Index estimations at EU level are available at the following website: http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_population_trends_of_farmland_birds

***MS and regional level* (examples)**: In Hungary the Common Bird Monitoring Programme has been running since 1999. A database of approximately 300 2.5x2.5 km survey squares is available for the whole timescale at the Monitoring Center of BirdLife Hungary. Detailed descriptions can be found at the following website (in Hungarian): http://www.mme.hu/mindennapi-madaraink-monitoringja-mmm

Fact sheets and information available from other sources: The proposed indicator is an alternative application of the Farmland Bird Index data sources; no direct fact sheet is available.

3. <u>Context of the case study testing</u>

Case study area: Heves Plain High Nature Value Area , Hungary

Policy context: Biodiversity decline is well-known throughout Europe, with agricultural habitats facing significant challenges. Pillar 2 measures can contribute to halt this overall decline. As several scientific studies and the programme evaluations show, well-targeted agri-environmental measures may hinder the further decline in agricultural biodiversity. Heves Plain High Nature Value Area is one of the most successful HNV area in terms of the uptake of the AE measures, thus provides a good opportunity for comparing the biodiversity values of participant and non-participant survey points. As the landscape is scattered by mosaic-like natural habitats (grasslands, wetlands, etc.) during the case study testing naturalness of the areas was also taken into consideration.

Used data:

1. Biodiversity data for 19 survey squares have been used for the last programming period of agri-environmental measures (2009-2014). In each monitoring square, data for 15 survey points are available, representing micro-level data for the exact location and within a 100m radius of the survey.
2. Participation data of the agri-environmental measures were used based on the Land Parcel Identification System.
3. Naturalness of the areas under examination was assessed by using CORINE 1:50 000 land cover data base.

Evaluation approach tested: A number of farmland bird individuals of the 285 survey points were compared based on the detailed grouping of the available data sets. Group design was based on the AE measure participation and the 'naturalness' of the survey points (share of the participant area inside the survey point/share of natural areas inside the survey point). Participant-natural, participant-non-natural, non-participant – natural, non-participant – non-natural groups were created, where the number of farmland bird individuals was assessed in parallel at the above mentioned time scale. Group design was carried by using spatial analyses tools (Jenks Natural Breaks method).

Table 24 Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Survey spots representing micro level by observing biodiversity data at parcel level.	Less frequently updated land-use data can limit the application of the indicator in elaborate statistics-based evaluations and result in unobserved impacts at local level	By using baseline data of a widely-known indicator, the proposed indicator may contribute to easier analyses of micro-level RD measure impacts.
Appropriateness of indicator(s) to capture complexity of environmental relationships	As the biodiversity data used is available as the baseline data of the Farmland Bird Index, data quantity limitations are not expected. Large sample size enhances the possibility of using multiple comparison groups as well as elaborate statistics-based methods to filter out other intervening factors.	Number of farmland bird individuals as an overall biodiversity indicator shall be further developed as this is rather sensitive to the effects of different years (weather conditions, migration circumstances, etc.).	As the data sources used are available in most of the EU Member States, the approach has high potential in replicability. More robust counterfactual assessment at micro level possible compared to using the FBI.

4. Recommended application

The use of the indicator of 'Number of farmland bird individuals' is recommended in cases where micro-level impacts of the different RD measures shall be detected, but biodiversity data gaps are observed at parcel level and the FBI cannot be used.

Baseline data of Common Birds Monitoring Programme shall be available, which means that the cooperation with the relevant monitoring organisations is highly recommended.

5.6.2 Indicator: Number of singing corncrake males

1. **<u>Definition / description of the indicator, including:</u>**

Environmental public good: Biodiversity Wildlife

Type of indicator: Additional (measure specific) indicator

Reflected RDP priority and focus area:

- Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
- Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.

Unit of measurement: Number of singing corncrake males

Type of data required: Regularly collected data on singing males of corncrakes, land cover data and agricultural land-use data (IACS).

Scale and level of application: The indicator is tested at micro level, scale of sampling plot – 0.28 km^2 (observation radius – 300 m).

2. **<u>Existing data sources</u>**

EU level: corncrake singing males census data is not systemically gathered, but is available in the countries which report about conservation status of corncrake according to the reporting requirements for EU Birds Directive implementation. Other necessary data, such as land-cover data (CORINE land cover or national equivalents), agricultural land-use data (IACS) is typically available.

***MS and regional level* (examples)**: In Lithuania corncrake census data is gathered within the framework of the state biodiversity monitoring programme. The monitoring is performed in the Natura 2000 areas designated for the conservation of this species.

Fact sheets available from other sources: no direct fact sheet is available.

3. **<u>Context of the case study testing</u>**

Case study area: Nemunas delta regional park, Šilutė municipality district, Lithuania.

Policy context: The indicator focuses on the impact of RDP Measure 214, which is one of the key measures to address biodiversity decline in grasslands under the CAP policy. In addition, the evaluation context directly relates to the EU biodiversity strategy implementation, in particular, target 3 “increase contribution of agriculture and forestry to biodiversity”.

Used data:

1. Corncrake density data are used collected from the national state biodiversity monitoring programme. There were 115 observation sample plots (circular form, radius 300 m) included in the evaluation.
2. Georeferenced spatial data set at a scale of 1:10,000 in the Republic of Lithuania (GDR10LT), Orthophoto images 2010-2014
3. Integrated Administration and Control System (IACS) available in GIS format for period of 2010-2013.
4. Additional databases were used: forest cadastre databases, Corine Land Cover 1:50,000. These databases were used for species environment analysis to determine side effects.

Evaluation approach tested*: testing focussed on defining causal linkages between the* occurrence of corncrakes and RDP measures, using data generated by the biodiversity monitoring programme carried out on a regular basis by the public sector. This involved taking evaluation steps through the developed logic model and analysing evaluation results eliminating various factors possibly impacting on evaluation conclusion. Functional unit for the corncrake was selected also on circular shape (diameter 600 m, covering 0.28 km^2) corresponding with standard observation point area as defined in corncrake monitoring methodology. Functional units were grouped according to the participation rate of the evaluated AEM within the sample areas. The functional units were grouped into multiple comparison groups according to intensity of participation in the targeted measure. The counterfactual scenario was defined based on with and without involving observation plots without participation of targeted measure, but with similar natural conditions. Multiple regression analysis was applied to analyse dependency of the corncrake numbers in the observation plot with different participation intensity. Results of the case study illustrate a robust statistical relation (p=0.01) between the number of observed birds and the participation intensity of Measure 214 in the functional unit. More corncrakes were observed in the fields where participation rate is higher. Different density rates cannot be explained solely by mowing activities, as monitoring was done in June when some of the non-participants had not yet mowed their areas. The higher density can be partly explained by the different (more natural) structure of the vegetation caused by participation in Measure 214, which leads to higher availability of food. However, if the monitoring would be performed later (or even better, a third count enabling comparison of the changes), results would be more informative indicating a clearer discrepancy between participants and non-participants.

Table 25 Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	The indicator can deliver robust data at micro level considering specific local characteristics of the assessed areas, which allows the evaluation at micro level	The indicator can be performed only for specific sub-measures within Measure 214 linked to limited types of grassland.	The indicator would contribute to more comprehensive evaluation as additional indicator together with FBI evaluation.
Appropriateness of indicator(s) to capture complexity of environmental relationships	Reflects robust causal relationships of biodiversity impacts of specifically targeted sub-measures of AEMs	The indicator is linked to narrow aspect of environmental problem/public good. The indicator shall be used in combination with other indicators e.g. FBI For a wider application of the indicator and an application with robust statistical counterfactuals adjustments to the corncrake census data methodology are needed (e.g. set later timing for the second count of birds).	The indicator provides an example for additional result indicators needed to assess biodiversity impacts of specifically targeted sub-measures of AEMs.

4. Recommended application

Number of singing corncrake males is recommended for the evaluation of impacts of Measure 214 at micro level. It could be relevant to use this indicator as an additional one along with FBI evaluation. Such an approach would contribute to FBI by providing additional information at micro level, while FBI itself is more a macro-level indicator.

More adjustments are needed in adapting the data gathering methodology for RDP evaluation usage. In addition, more research is needed to evaluate corncrake breeding success (and determine best timing for mowing) for this species to be an 'umbrella indicator'.

The indicator provides a good example of collaboration and data sharing potential between Agriculture and Environment sectors. There might be other data gathered by environmental sector, which could be successfully used for environmental evaluations of RDP measures.

It would be beneficial to have more coordination between environmental and agriculture authorities to determine data-sharing mechanisms. With adjustment, data could be gathered within the framework of the existing state biodiversity monitoring programme, leading to no additional costs (or a comparatively insignificant increase due to e.g. additional counts).

At the moment, such data is gathered only within designated Natura 2000 sites. It would be relevant to select a statistically robust number of samples outside protected areas, which would enable modelling results on macro level and national scale.

5.7 Animal welfare

5.7.1 Indicator: Animal-based / result-based indicators: Lameness and mortality rates

1. <u>Definition / description of the indicator, including:</u>

Public good: Animal welfare, Animal welfare category: Good health

Type of indicator: Programme-specific result indicators. No common result or impact indicators exist and managing authorities and evaluators are not formally required to define additional result or impact indicators targeted at animal welfare. However, for programmes which have implemented measures targeted at animal welfare (e.g. Measure 215 and 121), it is necessary to define and select suitable indicators to assess the effects of those measures on animal welfare. The sole use of output indicators is not sufficient. Animal-based indicators integrate a direct and result-based approach into the evaluation of animal welfare impacts. Lameness and mortality rates of cows and calves are two of the direct, i.e. animal-based, indicators to measure changes in the animal welfare category 'good health', established in the Welfare Quality® protocol. The indicator lameness measures changes in the share of lame animals compared to the total number of animals, while the indicator mortality rates measures the share of dead animals. The indicator lameness has causal linkages with policy measures targeted at improving housing conditions, such as type of bedding and the provision of straw, the provision of access to grazing and improving health care plans. The indicator mortality rates has causal linkages with policy measures targeted at improved feeding and water access, improving housing conditions such as type of beddings and space allowances and improving health care plans.

Reflected RDP priority and focus area: No focus area is particularly defined in relation to animal welfare in the CMES, but animal welfare is included in the rural development priority 3 "Promoting food chain organisation, including processing and marketing of agricultural products, animal welfare and risk management in agriculture".
Unit of measurements: Share of lame animals and share of died animals
Type of data required:

- Type of data required to measure / quantify the indicator:
    - Animal-based data: Livestock monitoring data either available from secondary data sources (e.g. the HIT database in Germany) or from empirical monitoring efforts through farm visits
- To enable the application of the indicators in an assessment of net-effects with advanced evaluation methods additional type of data are required:
    - Livestock husbandry and farm structural data: Data on husbandry systems and farm structural characteristics available from secondary data sources (e.g. FADN, Census data etc.)
    - Policy related data: IACS data on uptake of relevant measures (IACS database).

Scale and level of application: Farm level

2. Existing data sources

EU-level: No EU-wide data sources exist.

MS and regional level (examples): The indicators are collected as part of the benchmarking system of the Animal Health and Welfare Management Programme in Scotland (Measure 215) and are included in the HIT database in Germany.

Fact sheets and information available from other sources: No particular fact sheets exist, but Welfare Quality® Assessment Protocols have been developed by the Welfare Quality® consortium (2009).

3. Context of the case study testing

Case study area: North Rhine Westphalia, Germany

Policy context: Animal welfare payments - Measure 215, 2. Investment support with animal welfare related objectives – Measure 121.

Used data: Empirical monitoring data from farm visits (winter 2013 / 2014), HIT database.

Approach applied to review the suitability of the indicator (short explanation of the main logic model steps): The testing focussed on the development of guidelines for the selection of animal welfare indicators for RDP evaluation covering different relevant animal welfare criteria. Following the identification of the most relevant (and practical) animal welfare criteria which need and can be covered by the evaluation, different types of animal welfare indicators were reviewed and tested to inform the development of the guidelines for indicator selection. The case study differentiated between indirect indicators such as in relation to management and housing, and direct indicators such as in relation to animal health. Based on a review of stakeholder acceptance and practical feasibility of direct animal-based indicators (Bergschmidt et al., 2014 and 2015) advantages and disadvantages of different indicator types and the strengths and weaknesses of the integration of different result-based indicators in the animal welfare assessment were derived.

Table 26 Strengths and weaknesses of animal-based indicators (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Appropriateness of indicator(s) to capture complexity of animal welfare relationships [Lack of suitable animal welfare indicators in RDP evaluations]	Adds a direct (i.e. result-based) assessment of health criteria to the assessment of housing and feeding (water) criteria through the use of resource or management based indicators High acceptance by stakeholders and scientists Cost-effective application in combination with resource and management-based indicators feasible	Single indicator limited health aspects Cost-effective application depends on available monitoring data. High monitoring requirements and costs might prohibit the application if no data sources exist. Indicator can be influenced by seasonality	Improves the coverage of animal welfare impacts and contributes to a conceptually sound multi-criteria assessment of animal welfare

4. Recommended application

The application of the indicator is lameness recommended for the evaluation of animal welfare payments and investment support with an intervention logic linked to health and housing animal welfare criteria (micro level). The indicator mortality rates of cows and calves is best used in a multi-criteria assessment in combination with indicators on grazing access, increased space allowance and walking surface (resource and management-based indicators). The application of the indicator is recommended for the evaluation of animal welfare payments and investment support with an intervention logic linked to feeding (water), health and housing animal welfare criteria (micro level). The indicator is best used in a multi-criteria assessment in combination with indicators on feeding and water access, type of beddings and space allowances, and walking surface (resource and management based indicators).

The farm visits and livestock monitoring conducted in a pilot project by the Thünen Institute (Bergschmidt et al., 2015) highlight the high amount of staff resources required to monitor a sufficiently large sample in different years for RDP evaluations of animal welfare impacts. This implies that the feasibility of using these indicators in RDP evaluations depends on the availability of already existing monitoring data or secondary data sources. In case of long-term evaluation contracts, different sampling strategies can be explored to collect primary data through farm visits.

Practitioners and farmers 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 (Bergschmidt et al., 2015).

6. References

Action Plan for Thessaly Plain (2001) Join Ministerial Decision No 25638/2905 (Official Journal Government of Greece, 1422B)

Agriconsulting (2013) Aggiornamento Rapporto di Valutazione Intermedia del PSR 2007-2013, Regione Emilia Romagna, Bologna.

Alho J, Kolehmainen O, Leskinen P (2001) Regression methods for pairwise comparison data, the analytic hierarchy process in natural resource and environmental decision making. *Managing Forest Ecosystems* 3: 235-251

Arriaza M, Cañas-Ortega J, Cañas-Madueño J, Ruiz-Aviles P (2004) Assessing the visual quality of rural landscapes. *Landscape and Urban Planning* 69(1): 115-125.

Arthur L (1977) Predicting scenic beauty of forest environments: some empirical tests. *Forest Science* 23(2): 151-160.

Berg A, Gustafson T (2007) Meadow management and occurrence of corncrake Crex crex, Iris pseudacorus, Phragmites australis and Phalaris arundinacea. *Agriculture, Ecosystems and Environment* 120: 139–144

Bergschmidt A, Renziehausen C, Brinkmann J, March S (2014) Tiergerechtheit landwirtschaftlicher Nutztierhaltung: Verbesserung durch ergebnisorientierte Honorierung? *Ländl Raum* (ASG) 65(2): 32-33.

Bergschmidt A, Renziehausen C, March S, Brinkmann J (2015) Tierschutzwirkungen der Entwicklungsprogramme für den ländlichen Raum - Ergebnisse der Evaluierung der Maßnahme "Förderung umwelt- und tiergerechter Haltungsverfahren". In: KTBL, Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V. (Hrsg.): Herausforderung Tierwohl, KTBL-Tage: 208-223.

Belényesi M, Podmaniczky L (2007) A „Magas Természeti Értékű” mezőgazdasági területek lehatárolása. *Magyarországon Tájökológiai Lapok* 5 2: 347-362.

Bochu J-., Metayer N (2013) Development of Carbon Calculator to promote low carbon farming practices – User guidance manual for the Carbon Calculator, Deliverable to EC-JRC-IES by Solagro.

Botequilha Leitao A, Miller J. Ahern J, McGarigal K (2006) *Measuring Landscapes – a planner’s handbook*. Island Press, London.

British Standard Institute (2011) The Guide to PAS 2050:2011. How to carbon footprint your products, identify hotspots and reduce emissions in your supply chain, British Standards, London, UK.

Budka M, Osiejuk T (2013) Habitat preferences of Corncrake (Crex crex) males in agricultural meadows. *Agriculture, Ecosystems and Environment* 171: 33–38.

Carleer A, Wolff E (2006) Urban land cover multi‐level region‐based classification of VHR data by selecting relevant features. *International Journal of Remote Sensing* 27(6): 1035-1051.

Clay G, Daniel T (2000) Scenic landscape assessment: the effects of land management jurisdiction on public perception of scenic beauty. *Landscape Urban Planning* 49(1-2): 1-13.

Common Monitoring and Evaluation Framework (CMEF), Guidance note H – Output Indicator Fiches (http://ec.europa.eu/agriculture/rurdev/eval/guidance/note_h_en.pdf)

Common Monitoring and Evaluation Framework (CMEF), Guidance note I – Result Indicator Fiches (http://ec.europa.eu/agriculture/rurdev/eval/guidance/note_i_en.pdf)

Council of Europe (2000) European Landscape Convention, Council of Europe, Florence, October 2000.

Council Regulation (EC) No 1257/1999 of 17 May 1999 in support for rural development from the European Agricultural Guidance and Guarantee Fund (EAGGF) and amending and repealing certain Regulations (Official Journal L 160 , 26/06/1999)

Council Regulation (EC) No 1405/2006 of 18 September 2006 laying down specific measures for agriculture in favour of the smaller Aegean islands and amending Regulation (EC) No 1782/2003 (Official Journal L 265, 26/09/2006)

Council Regulation (EEC) No 2019/93 of 19 July 1993 introducing specific measures for the smaller Aegean islands concerning certain agricultural products (Official Journal L 184, 27.7.1993)

Cucek L, Kleme J, Kravanja Z (2012) A review of footprint analysis tools for monitoring impacts on sustainability. *Journal of Cleaner Production* 34.

Danezis I (2001) *Santorini: Thira, Thirasia, Aspronisi, Volcanoes*. Adam Publications, Athens

Daniel T (2001) Whither scenic beauty? Visual landscape quality assessment in the 21st century. *Landscape and Urban Planning* 54(1-4): 267-281.

de la Fuente de Val G, Atauri J, de Lucio J (2006) Relationships between landscape visual attributes and spatial pattern indices: A test study in Mediterranean-climate landscapes, *Landscape and Urban Planning* 77(4): 393-407.

Drosou K (2005) Present and Future of viniculture in Santorini, Management and Perspectives, Thesis for postgraduate studies in Agriculture and Environment, University of the Aegean.

Economou A, Skouteri A, Michopoulos P (2007) *Soils and Land Use of Santorini, Greece, Soils of Volcanic Regions in Europe*. Springer: 623-628pp.

European Commission (2013) Impact Indicators, Draft – Work in Progress for Discussion in the expert group on monitoring and evaluating the CAP. http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetailDoc&id=6707

European Environment Agency (2006) Report on Integration of environment into EU agriculture policy - the IRENA indicator-based assessment report. http://www.eea.europa.eu/publications/eea_report_2006_2/download

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

European Evaluation Network for Rural Development (2010) Working Paper on Approaches for assessing the impacts of the Rural Development Programmes in the context of multiple intervening factors (http://enrd.ec.europa.eu/enrd-static/fms/pdf/EB43A527-C292-F36C-FC51-9EA5B47CEDAE.pdf)

ENRD (2008) Guidance Document to the Member States on the Application of the HNV Impact Indicator. European Evaluation Network for Rural Development, Brussels.

Eurostat, Water abstraction. http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_water_abstraction

Eurostat, Gross Nitrogen Balance. http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_gross_nitrogen_balance

Fagan WF, Fortin MJ, Soykan C (2003) Integrating edge detection and dynamic modeling in quantitative analyses of ecological boundaries. *Bioscience* 53:730-738.

Farina A (2007) *Principles and Methods in Landscape Ecology – towards a science of landscape*. Springer Landscape Series 3.

Fonji SF, Taff GN (2014) Using satellite data to monitor land-use land-cover change in North-eastern Latvia. *SpringerPlus* 3: 61.

García-Feced C. Weissteiner CJ. Baraldi A. Paracchini M.L. Maes J. Zulian G. Kempen M. Elbersen B. Pérez-Soba M. (2015) Semi-natural vegetation in agricultural land: European map and links to ecosystem service supply, Agronomy for Sustainable Development, Vol. 35, Issue 1, pp 273-283.

Greece Then and Now: Longitudinal Mapping of Land Cover, 1987–2007, WWF Greece. http://www.wwf.gr/en/areas/forests/land-uses

Green RE (1996) Factors affecting the population density of the corncrake Crex crex in Britain and Ireland. Royal Society for the Protection of Birds, 17 Regent Terrace, Edinburgh, EH7 5BN, UK

Gregory R, van Strien A, Vorisek P, Gmelig Meyling A, Noble D, Foppen R, Gibbons D (2005) Developing indicators for European birds. *Philosophical Transactions of the Royal Society B* 360: 269–288.

Hansen AJ, di Castri F, Naiman R (1988) Ecotones: what and why? *Biology International* special issue 17 [A new look at ecotones: Emerging international projects on landscape boundaries]: 9-46.

Hofmann H.C., Rowe G.C., Türk A.H. (2011): Administrative Law and Policy of the European Union. p 561. Oxford University Press 2011

Howley P, Donoghue C, Hynes S (2012) Exploring public preferences for traditional farming landscapes. *Landscape and Urban Planning* 104(1): 66-74.

Hu Q, Wu W, Xia T, Yu Q, Yang P, Li Z, Song Q (2013) Exploring the use of Google Earth imagery and object-based methods in land use/cover mapping. *Remote Sensing* 5(11): 6026-6042.

IRENA Indicator fact sheet 18.1 Gross Nitrogen Balance. https://circabc.europa.eu/sd/a/09cc1348-8232-447e-ae03-87e71d680b08/IRENA%2018.1%20-%20Gross%20nitrogen%20balance.pdf

IRENA, Indicator Fact Sheet 34.3 Share of agriculture in water use. http://ec.europa.eu/eurostat/documents/2393397/2518916/IRENA+IFS+34.3+-+Share+of+agriculture+in+water+use_FINAL.pdf/78932951-7252-4de2-8998-caab964c4246

IRENA Indicator Fact Sheet 22- Water abstraction. http://ec.europa.eu/eurostat/documents/2393397/2518916/IRENA+IFS+22.pdf/c600296a-d688-446a-8d71-7f5fa05483d3

Johnson R, Brunson M, Kimura T (1994) Using image-capture technology to assess scenic value at the urban/forest interface: a case study. *Journal of Environmental Management* 40(2): 183-195.

Keenleyside, C, Beaufoy, G, Tucker, G, and Jones, G (2014) High Nature Value farming throughout EU-27 and its financial support under the CAP. Report Prepared for DG Environment, Contract No ENV B.1/ETU/2012/0035, Institute for European Environmental Policy, London.

Keišš O (2005) Impact of changes in agricultural land use on the Corncrake *Crex crex* population in Latvia. *Acta Universitatis Latviensis* 691: 93–109.

Kourakou-Dragona S (1994) *The Santorini of Santorini*. The Boutari Foundation, Athens.

Land cover and land use (LUCAS) statistics. http://ec.europa.eu/eurostat/statistics-explained/index.php/Land_cover_and_land_use_%28LUCAS%29_statistics

Likert R (1932) A technique for the measurement of attitudes. *Archives of Psychology* 140: Columbia University Press.

Lothian A (1999) Landscape and the philosophy of aesthetics: is landscape quality inherent in the landscape or in the eye of the beholder? *Landscape and Urban Planning* 44(4): 177-198.

McGarigal K (2015) Fragstats Help (http://www.umass.edu/landeco/research/fragstats/documents/fragstats.help.4.2.pdf)

Miller DR (2001) A method for estimating changes in the visibility of land cover. Landscape and Urban Planning 54: 93-106.

Minx JC, Wiedmann T, Wood R, Peters GP, Lenzen M, Owen A, Scott K, Barrett J, Hubacek K, Baiocchi G, Paul A, Dawkins E, Briggs J, Guan D, Suh S, Ackerman F (2009) Input–Output analysis and carbon footprinting: an overview of applications. *Economic Systems Research* 21(3): 187-216.

MMM (2007) The Rural Development Programme of continental Finland 2007-2013, approved 10.08.2007 (Manner-Suomen maaseudun kehittämisohjelma 2007 – 2013, Hyväksytty 10.8). In Finnish. http://www.maaseutu.fi/attachments/6BQPIuj8V/Manner-Suomen_maaseudun_kehittamisohjelma_2007-2013_090114.pdf. 375 p.

Morgan J, Gergel S, Coops N (2010) Aerial photography: a rapidly evolving tool for ecological management. *BioScience* 60(1): 47-59.

Nitrogen outputs (kg N per ha), 1990-2008, EU-27, CH and NO. http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Nitrogen_outputs_%28kg_N_per_ha%29,_1990-2008,_EU-27,_CH_and_NO.png

NLWKN (2015a) Anwenderhandbuch für die Zusatzberatung Wasserschutz Grundwasserschutzorientierte Bewirtschaftungsmaßnahmen in der Landwirtschaft und Methoden zu ihrer Erfolgskontrolle. Grundwasser, Band 21.

NLWKN (2015b) Erfolgskontrolle von Grundwasserschutzmaßnahmen mit Hoftorbilanzen eines Referenzbetriebsnetzes außerhalb der Trinkwassergewinnungsgebiete und der WRRL-Beratungskulisse. Grundwasser, Band 25.

NLWKN (2015c) Anwenderhandbuch für die Zusatzberatung Wasserschutz Grundwasserschutzorientierte Bewirtschaftungsmaßnahmen in der Landwirtschaft und Methoden zu ihrer Erfolgskontrolle. Grundwasser, Band 21.

NLWKN (2010) Untersuchung des mineralischen Stickstoffs im Boden. Empfehlungen zur Nutzung der Herbst-Nmin-Methode für die Erfolgskontrolle und zur Prognose der Sickerwassergüte. Grundwasser, Band 8.

Ode Å, Fry G, Tveit M, Messager P, Miller D (2009) Indicators of perceived naturalness as drivers of landscape preference. *Journal of Environmental Management* 90(1): 375-383.

Pandey D, Agrawal M (2014) Carbon footprint estimation in the agriculture sector. In: S Muthu, Assessment of Carbon Footprint in Different Industrial Sectors, Vol. 1, Springer, Singapore.

Paracchini M-L, Petersen J-E, Hoogeveen Y, Bamps C, Burfield I, van Swaay C et al. (2009) "Identification of High Nature Value Farmland at the EU27 Level on the Basis of Land Cover and Biodiversity Data", Joint Research Centre, Ispra (IT).

Pérez J (2002) Ascertaining landscape perceptions and preferences with pair-wise photographs: planning rural tourism in Extremadura, Spain. *Landscape Research* 27(3): 297-308.

Peters GP (2010) Carbon footprints and embodied carbon at multiple scales. *Current Opinion in Environmental Sustainability* 2: 245–250.

Purvis G, Louwagie G, Northey G, Mortimer S, Park J, Mauchline A, Finn J, Primdahl J, Vejre H, Vesterager JP, Knickel K, Kasperczyk N, Balazs K, Vlahos G, Christopoulos S, Peltola J (2009) Conceptual development of a harmonized method for tracking change and evaluating policy in the agri-environment: the Agri-Environmental Footprint Index. *Environmental Science & Policy* 12.

Rees, W. E. (1992) Ecological footprints and appropriated carrying capacity: what urban economics leaves out. Environment and Urbanisation 4 (2): 121–130.

Report of the Ex-post evaluation of the 2000-2006 Greek Rural Development Programme, (http://ec.europa.eu/agriculture/rur/countries/el/index_en.htm)

Report of the Mid-term evaluation of the 2007-2013 Rural Development Programme, (http://ec.europa.eu/agriculture/rurdev/countries/el/index_en.htm)

Reif M, Piercy C, Jarvis J, Sabol B, Macon C, Loyd R, Colarusso P, Dierssen H, Aitken J (2012) Ground Truth Sampling to Support Remote Sensing Research and Development: Submersed Aquatic Vegetation Species Discrimination Using an Airborne Hyperspectral/Lidar System, ERDC TN-DOER-E30 (www.coastalamericafoundation.org)

Rousou A (2006) Conventional and organic farming on Santorini's vineyard, Thesis for the department of Crop Science, School of Agricultural technology, Technological Educational Institute of Crete.

Rural Development Programme of Greece 2007-2013, Hellenic Ministry of Rural Development and Food (http://www.agrotikianaptixi.gr/Uploads/Files/13th_modification_27_6_2014.pdf)

Saaty T (1977) A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology* 15(3): 234-281.

Sang N, Ode A, Miller DR (2008) Landscape metrics and visual topology in the analysis of landscape preference. Environment and Planning B 35: 504-520.

Sarpaki A Jones G (1990) Ancient and modern cultivation of Lathyrus clymenum L. in the Greek Islands. *The Annual of the British School at Athens* 85: 363-368.

Sharp R, Tallis H, Ricketts T, Guerry A, Wood S, Chaplin-Kramer R, Nelson E, Ennaanay D, Wolny S, Olwero N, Vigerstol K, Pennington D, Mendoza G, Aukema J, Foster J, Forrest J, Cameron D, Arkema K, Lonsdorf E, Kennedy C, Verutes G, Kim CK, Guannel G, Papenfus M, Toft J, Marsik M, Bernhardt J, Griffin R, Glowinski K, Chaumont N, Perelman A, Lacayo M, Mandle L, Hamel P, Vogl A, Rogers L, Bierbower W (2015) InVEST +VERSION+ User's Guide. The Natural Capital Project, Stanford University, University of Minnesota, The Nature Conservancy, and World Wildlife Fund.

Szép T, Nagy K, Nagy Z, Halmos G (2012) Population trends of common breeding and wintering birds in Hungary, decline of long-distance migrant and farmland birds during 1999–2012. *Ornis Hungarica* 20(2): 13–63.

Szép T, Nagy K, Nagy Z (2015) Evaluation of the indicator data of the Agri-environmental Monitoring System. Internal report for the National Food Chain Safety Office.

Tahvanainen L, Ihalainen M, Hietala-Koivu R, Kolehmainen O, Tyrväinen L, Nousiainen I, Helenius J (2002) Measures of the EU Agri-Environmental Protection Scheme (GAEPS) and their impacts on the visual acceptability of Finnish agricultural landscapes. *Journal of Environmental Management* 66(3): 213-227.

Tahvanainen L., Tyrväinen L, Nousiainen I (1996) Effect of afforestation on the scenic value of rural landscape. *Scandinavian Journal of Forest Research* 11(1-4): 397-405.

Tuomisto H, Angileri V, De Camillis C, Loudjani P, Pelletier N, Nisini L, Haastrup P (2013) Final technical report: Certification of low carbon farming practices. JRC Technical Reports. European Commission Joint Research Centre. Luxembourg.

Tveit M, Ode Å, Fry G (2006) Key concepts in a framework for analysing visual landscape character. *Landscape Research* 31(3): 229-255.

Tyrväinen L, Tahvanainen L (1999) Using computer graphics for assessing the aesthetic value of large-scale rural landscapes. *Scandinavian Journal of Forest Research* 14(3): 282-288.

Tyrväinen L, Silvennoinen H, Kolehmainen O (2003) Ecological and aesthetic values in urban forest management. *Urban Forestry and Urban Greening* 1(3): 135-149.

Vitali A, Lo Presti S, Schipani T, Nardone A, Lacetera N (2014) Evaluation of regional agri-environmental policy on livestock greenhouse gas emissions. In: Proceedings of Livestock, Climate Change and Food Security Conference 19-20 May 2014 Madrid, Spain

Welfare Quality® consortium (2009) Welfare Quality® Assessment Protocol for Cattle ISBN/EAN 978-90-78240-04-4, 180 pages.

Wherrett J (2000) Creating landscape preference models using the Internet as a medium for surveys. *Landscape Research* 25(1): 79-96.

Wiens JA (1992) What is landscape ecology, really? *Landscape Ecology* 7:149-150.

http://www.thira.gr/dimos.html Municipality of Thira

www.thira.gr/epixeirisiako-programma.html, Operational Plan of Thira Municipality for 2013-2014

www.statistics.gr, National Statistical Service of Greece

http://www.peliti.gr/index.php?option=com_content&view=article&id=232:sandorini&catid=11:fitiko&Itemid=16