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: Development and application of methodologies and tools, including indicators, for the assessment of environmental impacts of rural development programmes in the EU

Report D4.3
Summary report on the methodological framework at micro level

---

Authors: Andrea Povellato, M. Valentina Lasorella, Davide Longhitano (CREA-INEA)

- with input from the AUA, BEF, JHI, LUKE, SZIE and TI teams

Approved by Work Package Manager of WP4: Andrea Povellato (CREA-INEA)

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

List of Figures

List of Acronyms

ABM	Agent Based Modelling
AEP	Agri-Environment Payments
CAP	Common Agricultural Policy
CGE	Computable General Equilibrium
CLC	Corine Land Cover
CMEF	Common Monitoring and Evaluation Framework
CMES	Common Monitoring and Evaluation System
DD	Difference-in-Difference
DR	Data Requirement
FADN	Farm Accountancy Data Network
FU	Functional Unit
GNB	Gross Nutrient Balance
HNV	High Nature Value
IACS	Integrated Administration and Control System
IV	Instrumental Variable
LCA	Life Cycle Analysis
MI	Methodological issues
NB	Naïve Baseline
NG	Naïve Group
PM	Pipeline Methods
PSM	Propensity Score Matching
PSM-DD	Propensity Score Matching Double Difference
QA	Qualitative Analysis
RD	Regression Discontinuity
RDD	Regression Discontinuity Design
RDP	Rural Development Programme
SE	Structural Econometric
SEEA	System of Environmental-Economic Accounts
SELU	Socio-Ecological Landscape Unit
SES	Socio-Ecological System

SNA	System of National Accounts

1 Summary

The environmental impact assessment of rural development policies is an issue that needs to be constantly monitored by both practitioners and policy makers within the EU and its Member States. Impact assessment methodologies are well established in the literature on agri-environmental policies, both at micro and macro scale, considering different public goods (water, biodiversity, etc.) as specific foci of these policies. However, there are several challenges and gaps related to their use, concerning both the fitness of indicators, models and methodologies for the expected outcomes, and the adoption of the most suitable scale for the analysis.

This report presents a conceptual framework in order to systematise the most commonly used methodologies to assess micro-level environmental performance of agricultural policies. It deals with issues relating to the specific use for some methodologies (or models, or indicators) for a single public good, aiming to contribute to systematise the current knowledge on evaluation methods. It tries to clarify the role of methodologies and their integration in the evaluation process, particularly to fill the gap of knowledge of the agriculture-environment relationships within the complexity of multi-scale and multi-levels approaches.

From a micro-level perspective, it is important to consider the role of individuals and analyse in depth the different forms of organisation (spatial, networks, hierarchies) and interactions among different organisational and intervening levels. Only with multi-scale integration and the combination of results it is possible to efficiently generalise (up-scale) micro-level results in a macro-level perspective. Field measurements, farm management surveys and farming system models essentially refer to the farm as the simplest management unit of an agricultural system, analysed from the point of view of a farmer who decides whether or not to participate in rural development schemes.

In most of the case study areas only naïve quantitative analysis has been applied due to difficulties in data availability and data access, with negative effects from the methodological point of view when the statistical significance of the parameters was not verified. The statistics-based approach to the counterfactual needs well-defined samples with a sufficient number of observations to perform regression models and spatial analysis.

2 Background

2.1 Rationale for the framework

The environmental impact assessment of rural development policies is an issue to be constantly monitored by both practitioners and policy makers within EU and its Member States. The empirical evaluation of the policy effects reveals, in fact, strengths and weaknesses of the applied strategies to enable better design, and outlines more responsive and effective policies for agri-environmental practices. Thus, impact assessment methodologies are well established in the literature on agri-environmental policies, both at micro and macro level, considering different public goods (water, biodiversity, etc.) as specific foci of these policies. In these terms, a vast range of micro-level methodologies is available for impact assessment and evaluation. However, there are several challenges and gaps related to their use, concerning both the fitness of indicators, models and methodologies for the expected outcomes, and the adoption of the most suitable scale and level for the analysis.

Starting from these overall considerations, this report tries to deal with these challenges, taking into account the experiences gained from past and current evaluations of RDPs and more generally from specific assessment of the relationships between agriculture and environment. Stakeholders involved in project workshops and meetings (evaluators, managing authorities and monitoring agencies above all) identified a number of gaps in evaluation methods, as also showed in scientific literature (Primdahl et al. 2010). Need for improved clarity in objectives; great variations in practices about the relevance and the use of indicators; lack of appropriate targeting approaches; insufficient or absent baseline data; different approaches to reporting and deficient evaluation frameworks (including lack of appropriate impact models) represent as much challenges for ENVIEVAL project.

The evaluation of RDP impact on the environment consists of three main components: a sound counterfactual design[1] , and assessments at micro and macro[2] levels. This report presents the micro level component of a conceptual framework that structures the current methodologies to assess micro-level environmental performance of agricultural policies, used in the scientific field and by practitioners in the policy evaluation assessment (Figure 1). It deals with issues related to the specific use of micro level methodologies, models and indicators compares their suitability for the evaluation of environmental impacts on a particular single public good. Future trends of evaluation methods should address new developments able both to integrate diversified approaches (for example qualitative and quantitative), and to consider different perspectives of environmental issues. Finally, with the aim of contributing to structure the current knowledge on evaluation methods, this report tries to clarify the role of methodologies and their integration in the evaluation process, particularly to fill the gap of knowledge of the agriculture-environment relationships within the complexity of multi-scale and multi-level approaches.

[1] For more detail see Artell et al. (2015) on the methodological framework for counterfactual development.
[2] For more detail see Aalders et al. (2015) on the theoretical and methodological framework for macro-level.

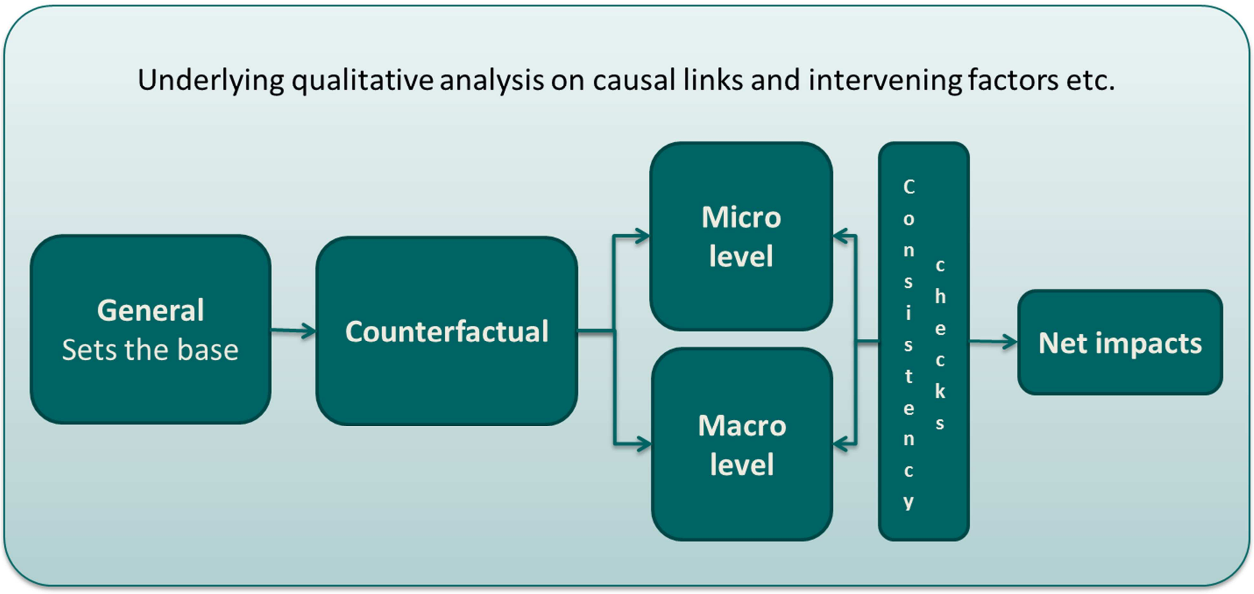

Figure 1 Simplified logic model flow of evaluation

2.2 Challenges

The micro level is often represented by the farm, which is considered as the simplest management unit of an agricultural system, analysed from the point of view of a farmer as the final decision-maker. In the case of RDP measures, the observation of changes related to the territorial focus prevails, and a common practice is to assess the impact at micro level and then scale up to macro level. The 'macro-system' is characterised by interacting components (micro level), where each component is part of the system (farming systems). Consequently, its behaviour does not only derive from the sum of effects, but from the presence of 'emergent properties' that occur only in a certain state of organisation. In this context it is important to consider the role of individuals and analyse in depth the different forms of organisation (spatial, networks, hierarchies) and interactions among different organisational and intervening levels.

The assessment needs to be carried out at an appropriate scale able to represent both natural processes and socio-economic systems, in order to include multiple benefits and potential for cumulative environmental impacts. The multi-scale integration and combination of results should provide the possibility to generalise efficiently micro-level results in a macro-level perspective. Efforts are required to explore the micro-level environmental effects in a macro perspective, and to exploit results obtained at the farm level to describe more general performances in agri-environmental schemes between micro and macro approaches.

For both challenges fit-for-purpose data, datasets and data sources are required for more appropriate and holistic analysis and evaluation. The lack of appropriate and specific data can undermine the results of the evaluation exercises. Furthermore, taking into account the current and past experience of RDP evaluations, the difficulties encountered by evaluators to use complex methodologies could weaken a good outcome from the evaluation process. The most commonly adopted approaches are based on sampling methods and/or integrated (biophysical) models. Both have some advantages and disadvantages in terms of generalisation of the primary micro-level findings to a different scale perspective. The adoption of these models needs specific datasets, for a vast range of socio-economic, environmental and institutional variables and long-term coverage for comparative analysis. Micro-level data should be developed in a more consistent and standardised way, targeting an accurate data collection at farm level, in order to provide a detailed overview of the whole farming system. An emerging question is related to the representativeness of the data collected at farm level. Finding methods that ensure data representativeness is crucial for future challenges.

The measurement of net impacts has to consider the indirect effects of the implementation of an environmental policy that, at micro level, could be: a) deadweight loss effects when changes occurred even without the measure implementation; b) leverage effects as inducing behaviour for other farmers in terms of practice changes. While the deadweight effects have a direct link with environmental impacts, the leverage effects are more concerned with the socio-economic side of the impact and are less relevant in terms of environmental impacts. The estimation of deadweight loss is particularly challenging in terms of: a) data available for the creation of the control groups without selection bias; and b) identification of appropriate indicators that synthesise the causal links between farm inputs and outputs and environmental outcomes.

2.3 Linkages between land management activities and environmental impacts

The environmental impact of agriculture depends to a large extent on the wide variety of agricultural practices adopted by the farmers. The way agriculture affects ecological systems is very complex due to multiple relationships between farming activity and environmental quality, depending on climate variables such as rainfall and temperature and on the physical conditions of the soil. Environmental impact involves a variety of factors from the decline in soil productivity to non-point source water pollutants, from overusing of surface and ground water for irrigation to the loss of wetlands and wildlife habitat and the reduction of genetic diversity. On the other hand some types of farming help to preserve habitat useful for wildlife species or create landscape highly appreciated from the socio-cultural point of view.

In order to assess the environmental impacts of farming activities both in quantitative and qualitative terms, a bulk of scientific literature is available, as presented in Deliverable 4.1 (Povellato et al., 2013) with specific reference to the evaluation of policy effects at micro level. Methodologies, models and indicators have to be framed and integrated to increase the knowledge of a generic evaluation process of human-environmental relationships. In this context, models and indicators play a crucial role by providing methods and tools for the assessment of agro-ecosystems and their environmental effects. Indicators represent the first functional component, able to monitor social and environmental developments at various temporal and spatial scales. They should be organised preferably within conceptual frameworks that help to logically organise the information, whilst models provide methods and tools to support the analysis of specific systems (in this case agro-ecosystems), or more generally the territorial systems in which agri-environmental issues are considered.

A system can be defined as the limited part of reality that contains a set of interacting or interdependent component parts and it is delimited by its spatial and temporal boundaries. A model is a simplified representation of a system, where simulations, built on mathematical models, allow their properties to be studied, in relation to those of the referred systems. Although a model always simplifies reality, it should contain all the essential features of the real system, in order to describe and solve problems. The balance among simplification, comprehensiveness and effectiveness depends upon the scope of the model, which may be very diverse, thus determining a wide range of possible model typologies (Giupponi and Carpani, 2006).

Indicators and models use available knowledge derived from specific forms of analysis to gain information and insight for a specific assessment purpose (e.g. assessment of environmental impact of the implementation of public policy). Furthermore, indicators play a fundamental role as a communication interface between science and policy decision-making and for communicating the performance of the farming systems-environment relationships in a concise and effective way, trying to bridge the gap between producers and information users, i.e. between the information available through scientific resources and the need for information for decision making at public and private level.

The environmental analysis of agricultural systems should first of all identify proper ways to describe the phenomena to be assessed. Significant measurable variables should be identified and processed to transform the acquired data into information. The environmental impact depends on the production practices of the system used by farmers. The connection between emissions into the environment and the farming system is largely indirect, due to the fact that emissions to the environment depend of the type of farming practice adopted and other random factors such as temperature, rainfall, etc (van der Werf and Petit, 2002).

2.4 Unit of analysis and unit of observation

Field measurements, farm management surveys and farming system models appear to be the key determinants for a good evaluation of the environmental impacts at micro level and the starting point for upscaling the outcomes at agroecosystem or landscape or regional levels (macro levels). In this context the choice of the suitable unit of analysis is crucial to conceptually join farming activities, environmental impacts and, as necessary, agricultural policies to be evaluated.

The unit of analysis is the most elementary part of the phenomenon to be analysed and its definition influences the design of analysis and data collection (Frankfort-Nachmias and Nachmias, 1996). The unit of observation represents the objects that are observed and about which information is systematically collected. More specifically, this can be synonymous with statistical units or sampling units. The observational unit is determined by the method by which observations have been selected. The unit of analysis is the object about which generalisations are made based on explaining a specific phenomenon.

The distinction between the two units is not always straightforward. A study may have a different unit of observation and unit of analysis when, for example, the research design may collect data at the individual level of observation (e.g. farm) but the level of analysis might be at an upper level (e.g. landscape), drawing conclusions from data collected from individuals. In some other cases the two units are the same when the generalisations being made from a statistical analysis are attributed to the unit of observation.

In the case of environmental assessment, the ecological relationships, referred to the unit of analysis, are very important to establish the appropriate links between land management and ecosystem elements. The definition of ecological unit as a distinct combination of landscape elements may be useful to evaluate the effects of management actions on ecosystems. Jax (2006) proposes to focus on process in comparison to that on statistical patterns, to choose the type of boundaries that defines ecological units and to establish the degree of internal relationship that is required to identify a specific unit. In the ecosystem approach, the emphasis is on the interactions between the components of the units (process-based or functional view) while the statistical approach is frequently used for mapping or classification of ecological units. The process-oriented approach is applied when predictions of ecological dynamics are intended. Boundaries of the ecological units are drawn according to either to discontinuity in space (topographical delimitation) or on the basis of the extension of functional relationships between the elements of the unit (functional delimitation).

Topographical boundaries are common in many definitions of ecological unit and the main challenge is represented by assessment of the homogeneity of the environment in space. Also for the functionally defined boundaries the most critical aspect is the assessment of homogeneity of process, which is a matter of scale and of the observation variables selected. The internal relationships of units can be measured as a gradient composed of different and specific degrees of interactions between the elements that are characterised by a specific role, working together as a cybernetic self-regulating system and losing their autonomy as part of the whole (Jax, 2006).

A well-established definition of the term 'functional unit' (FU) is used in the Life Cycle Analysis (LCA) context, where FU describes and quantifies those properties of the product which must be present for the studied context to take place. These properties (the functionality, appearance, stability, durability, ease of maintenance, etc.) are in turn determined by the requirements in the market in which the product need to be sold (Weidema et al., 2004). The reference flow is a quantified amount of product(s), including product parts, necessary for a specific product system to deliver, and translates the abstract FU into specific product flows for each of the compared systems, so that alternative product is compared on an equivalent basis, reflecting the actual consequences of the potential product substitution. The reference flows are the starting points to build the necessary models of the product systems (De Benedetto and Klemeš, 2009). The functional unit determines equivalence between systems and allows for comparison between them (Peters, 2010). In a comparative LCA, the functional unit shall be the same for all the compared product systems. This is a prerequisite for ensuring equivalence among the product systems (Weidema et al., 2004; Whittaker et al., 2013).

More in general, Functional Unit is a concept related to different disciplines and contexts. In an ecological context, 'function' is related to the structural components of an ecosystem (e.g. vegetation, water, soil, atmosphere and biota) and how they interact with each other, within ecosystems and across ecosystems. Sometimes, ecosystem functions can be found as a synonym of ecological processes and include stocks of materials (e.g., carbon, water, mineral nutrients) and rates of processes involving fluxes of energy and matter between trophic levels and the environment. In the Convention on Biological Diversity there is an explicit reference to FU in the adopted definition of an ecosystem as "*A dynamic complex of plant, animal, and microorganism communities and their non-living environment interacting as a functional unit*". The term FU has only been used for this definition in the conceptual framework of the Millennium Ecosystem Assessment (2003), without any additional reference to it. In the ecosystem approach adopted by the CBD, the definition of the basic structural and functional units of the ecosystem is considered in a pragmatic way. Depending on the scale, the guiding principle is that a well-defined ecosystem has strong interactions among its components and weak interactions across its boundaries (Smith and Maltiby, 2001).

Another interesting approach to ecological/functional units can be found in the System of Environmental-Economic Accounts (SEEA) which expands the scope of assets beyond the boundaries of the System of National Accounts (SNA). Accounts in physical units aim at supplementing conventional national accounts with data on the use and availability of natural resources. Accounting requires clear definition of the units of analysis for creating consistent databases supporting statistical comparison in space and time. In SNA the basic ('institutional') units are represented by legal entities that have the capacity to take any economic decision concerning production, consumption, investment, etc. In ecosystem accounts, equivalent units have to be defined moving from economic or administrative units to statistical units adapted at measuring environmental outputs and natural capital.

The European Environment Agency has proposed a simplified framework of ecosystem capital accounts based on the definition of socio-ecological system (SES) which integrates ecosystem functions and dynamics as well as human activities and the interaction of all these (Weber, 2011). The equivalence of SES with SNA's institutional unit allows the creation of consistent and integrated databases between economic and environmental accounts. In socio-ecological systems, natural and socio-economic elements interact to transform ecosystem functions in goods and services. However, it is important to distinguish between theoretical units, which help to describe the analytical framework and the observation units which are proxies that may be used for practical reasons to collect data (Weber, 2014). The basic statistical unit is the Socio-Ecological Landscape Unit (SELU) derived from the Corine land-cover maps and additional geo-environmental information on a 1 km grid. Within these landscape units, three groups of services (biomass/carbon production, freshwater production and functional services) are assessed, using respectively tonnes of carbon, cubic metres of water and, for the latter, very heterogeneous kind of ecosystem services, a composite index allows us to measure the capacity or potential of ecosystems to deliver ecosystem services in a sustainable way.

3 Micro level logic model

3.1 The three logic models for micro level

The workflow for the micro-level logic models leads to different methods which contribute, through the integration of micro and macro-level results, to a consistent net impact assessment. For each of the three possible counterfactual designs, an individual micro-level logic model has been created. The initial two phases of the workflow for these three logic models are the same, and it is only from the third phase that each of the counterfactual approaches leads to different micro-level methods which are discussed separately later.

Step 3.1 - Definition of the Unit of Analysis and Indicators

The micro logic model starts with the general layer on the data availability for all the three counterfactual approaches (Figure 2). To a large extent data availability determines the type of unit of analysis that can be used in the evaluation process and it provides information on the suitable indicators to be developed according to level and scale used in the analysis.

The selected indicators subsequently define the specific scale at micro level. Available data and linkages to micro-level results are considered for the identification of specific observational units, which leads to consistent indicators for the assessment.

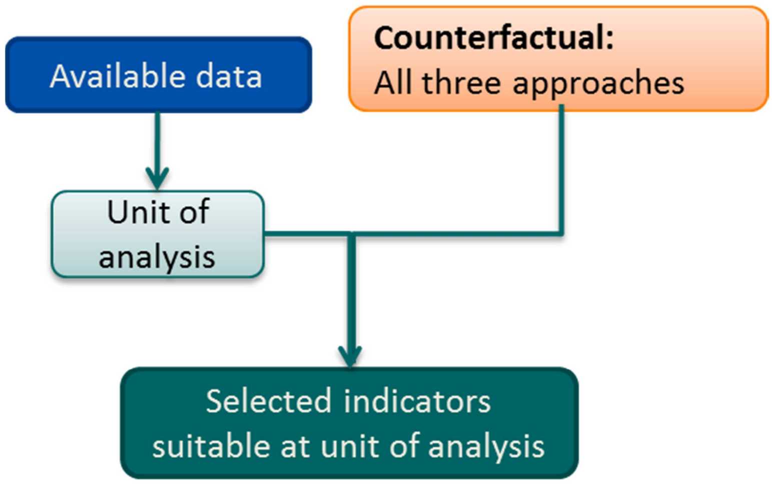

Figure 2 Definition of the Unit of Analysis, Unit of Observation and Indicators (Step 3.1 - micro level logic model)

Step 3.2 - Assessment of data quality

In the second step, assessments of the quantity and quality[3] of data have been carried out in order to check if the amount and characteristics of data are appropriate to implement one of the methods available for the impact evaluation (Figure 3). The limited data quality often affects the applicability of the methods for the environmental impacts assessment leading to a lack of consistent, robust and representative results. For this reason, an essential point in the workflow is the identification of potential bottlenecks, due to poor data quality, that can make the calculation of the selected indicators through one of the methods inadequate or require an increase in the quantity of data (e.g. number of observation) to assure a better representativeness of the results. Furthermore, in the case of environmental assessment, the availability of spatially-explicit data could make the difference between a rather descriptive survey and a more in-depth scientifically sound analysis. Starting from the selected indicators, a first check of the suitability of the data could require new primary data to be collected through statistical sampling. A second check could be needed to obtain sufficiently accurate data, possibly spatially explicit. In this case, additional data and/or particular data processing are required to improve data quality and quantity.

[3] Among the various categories of attributes of data quality, the most commonly attributes included are: accuracy, correctness, currency, completeness and relevance.

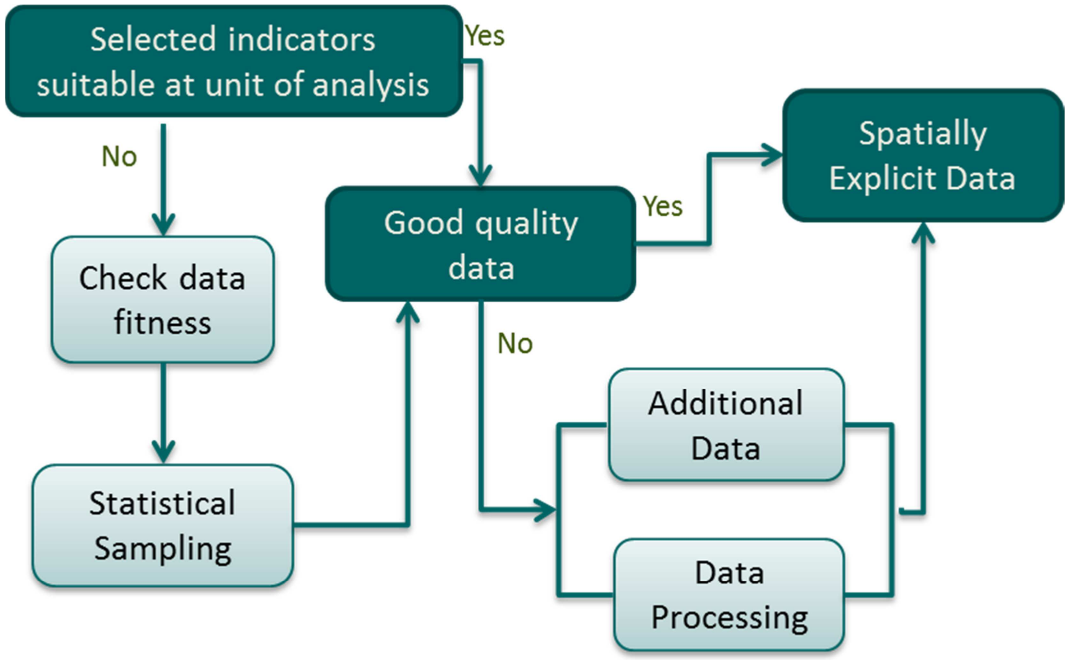

Figure 3 Assessment of data quality and quantity (Step 3.2 - micro level logic model)

Sometimes poor data quantity is also due to the lack of access to administrative and statistical databases because of privacy and data protection regulations or ill-coordinated efforts to collect information for monitoring purposes. The minimal requirements for suitable data sources should be causally linked to each other and frequently monitored. For this reason, the use of qualitative approaches (common sense) is quite frequent, not only for the lack of data or financial resources for creating new databases but also for the difficulties encountered by evaluators using complex methodologies that could guarantee a good outcome from the evaluation process. This knowledge gap has to be taken into account during the selection of a specific method which needs suitable data.

After the data are collated and verified, selected methods on the basis of the three counterfactual approaches have been identified.

3.2 Choice of methodology

Step 3.3a - Long Run 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 4).

In presence of the above categories, examples for selected methods are the structural model, integrated models and agent-based models, according to the availability of spatially explicit data. Without spatial data, *structural models* are more appropriate at the micro level. These models are defined by a mathematical approach to study the link between cause-effect relationships. More precisely, the method builds a framework for interpreting policy effects due to specific interrelationships among endogenous variables and exogenous variables or factors without the necessity of a comparison group. This allows capturing the effects of specific environmental policies at micro level, due to focus of cause-effect relationships. In general the structural model can be used to estimate unobserved or behavioural parameters.

Where there is availability of spatial data, the methods selected are the integrated models and the agent-based approaches. *Integrated models* allow agri-environment evaluation questions to be addressed more holistically, in particular at the farm scale and its sub-sets, such as cropped areas or parcels. In fact, this is the level for which farmers allocate available land and resources to the various tasks in their production systems. Integrated models are therefore able to shed light on the environmental components allowing evaluation of specific programmes. The environmental impacts of these changes can be estimated introducing linkages with bio-physical models at farm scale.

To date, researchers use farm-level decision models to assess behaviours and changes with *Agent-Based Modelling* (ABM) approaches in ex-ante evaluation exercises. These approaches allow the coupling of environmental models and the social systems embedded in them. In this way the role of social interactions of adaptive, disaggregated (micro-level) human decision-making processes in environmental management can be modelled. In short, the development and use of ABMs for ecosystem management allow consideration of ecological complexity. It is possible to identify the role of individuals and to analyse in more depth and more effectively the different forms of organisation (spatial, networks, hierarchies) and interactions among different organisational levels. However, the complexity of the ABM models and the intrinsic characteristics as simulation model mainly suggest using them on ex ante evaluation.

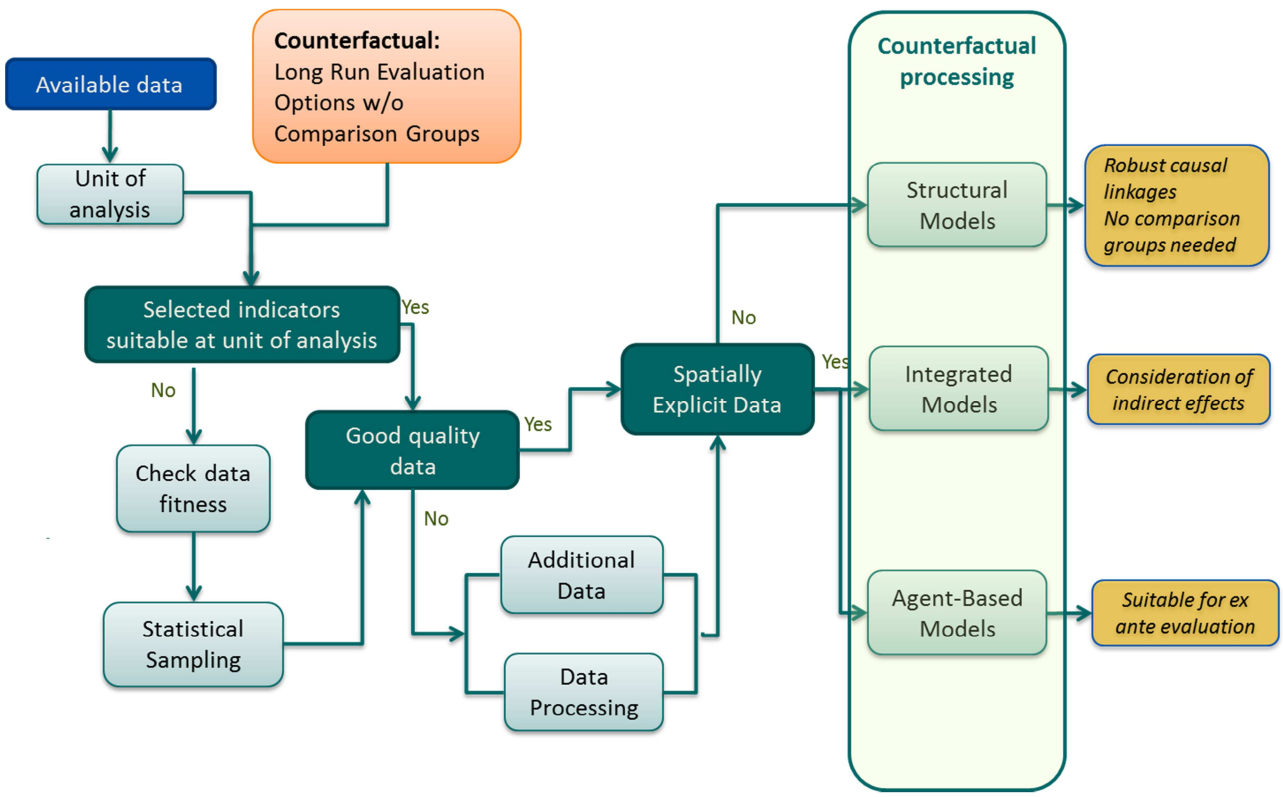

Figure 4 Long Run Evaluation Options w/o comparison groups (Step 3.3a - micro level logic model)

Step 3.3b - Naïve Estimates of Counterfactual

Naïve estimates of counterfactual should be used when data on programme participants prior and after programme are generally available, but not at a sufficient level of quality and quantity to use elaborate statistics-based approaches to assess net-effects at micro level. It can be divided in three different techniques: (i) the naïve 'before-after' estimator, which utilises programme data on programme participants to compute programme outcomes for programme participants (without counterfactual); and (ii) the naïve 'with-and-without' approach, that use the non-participants as a control group (Figure 5) and iii) the naïve application of a difference in difference approach.

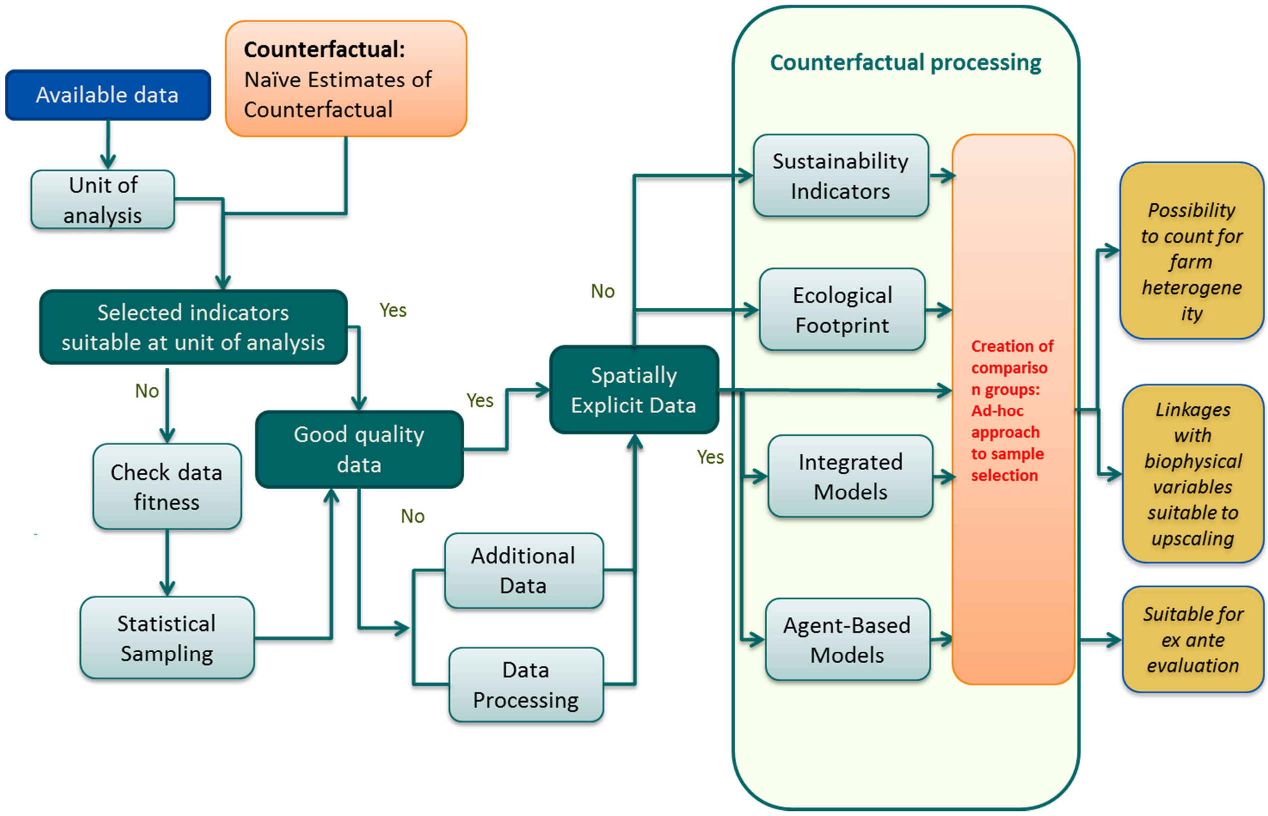

Figure 5 Naïve Estimated of Counterfactual (Step 3.3b - micro level logic model)

These approaches are based on the assumption that, in the absence of the programme, the outcome indicator of the participants in the programme would be the same as for non-participants in the same programme. The control group in the naïve comparison of programme participants is represented by the population average of non-participants. In this evaluation approach the data necessary for the average outcome indicators in the group of non-participants is usually obtained from statistical databases. Sometimes in this specific counterfactual design, there is no need for application of specific method to obtain the information necessary for the assessment, if sufficient self-explanatory variables are available. Otherwise there are some methods that can create the final 'indicator' adapted for the naïve and qualitative evaluation.

The possible methods linked with the naïve approaches at micro level are sustainability indicators, ecological footprint, integrated models and ABMs. In the first case, no spatially-explicit data are necessary. With the ecological footprint, and more in general with the use of composite sustainability indicators, it is possible to count the farm heterogeneity due to human environmental action to define better policy evaluations within a single agricultural system. In the case of availability of spatial data, integrated models and ABMs should be used, which have the characteristics described in the previous section. Basically the use of all these methods allows designing the counterfactual on the basis of the data commonly available from official statistical sources available at local level (e.g. FADN, Census, FSS).

Step 3.3c - Elaborate Statistics-based Evaluation Options

In this case four options can be defined:

i) comparing samples of participants and non-participants using matching approaches (e.g. propensity score matching) to compare groups of participants and non-participants with the same characteristics and propensity to participate
ii) conducting an intermediate counterfactual analysis between different participant groups (e.g. participants and late joiners);
iii) using similar non-eligible farms to represent non-participants (regression discontinuity method);
iv) comparing farms participating and those in queue together (pipeline method).

For this approach the abundant data availability about general characteristics and performance of participants and non-participants, before and after implementation of the RDP is essential. As for the naive evaluation approach, also in this case the application of 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: the Difference in Differences (DID); the regression discontinuity design (RDD); the matching methods and propensity score matching (PSM); and the combined methods. The DID compares the before and after changes of programme participant and after change of outcome indicators. This approach allows control of the unobserved heterogeneity (under the assumption that this does not vary in time). This method requires data availability between two periods observed (time series). The RDD requires availability of dataset with variable and observation on eligible and non-eligible units, with time series of cross-sectional data. In fact the RDD allows assessment of the effects of programmes that have a continuous eligibility. The matching methods, including the PSM, are the most advanced and effective tools of evaluation. They are based on advanced statistical approaches and need abundant data on participants and non-participants, requiring high quantitative skills of evaluator. Through this approach at the micro level, the methods selected are the ecological footprint and the integrated models. In the first case, no spatial explicit data are necessary on the contrary in the integrated models explicit data are necessary (figure 3.5).

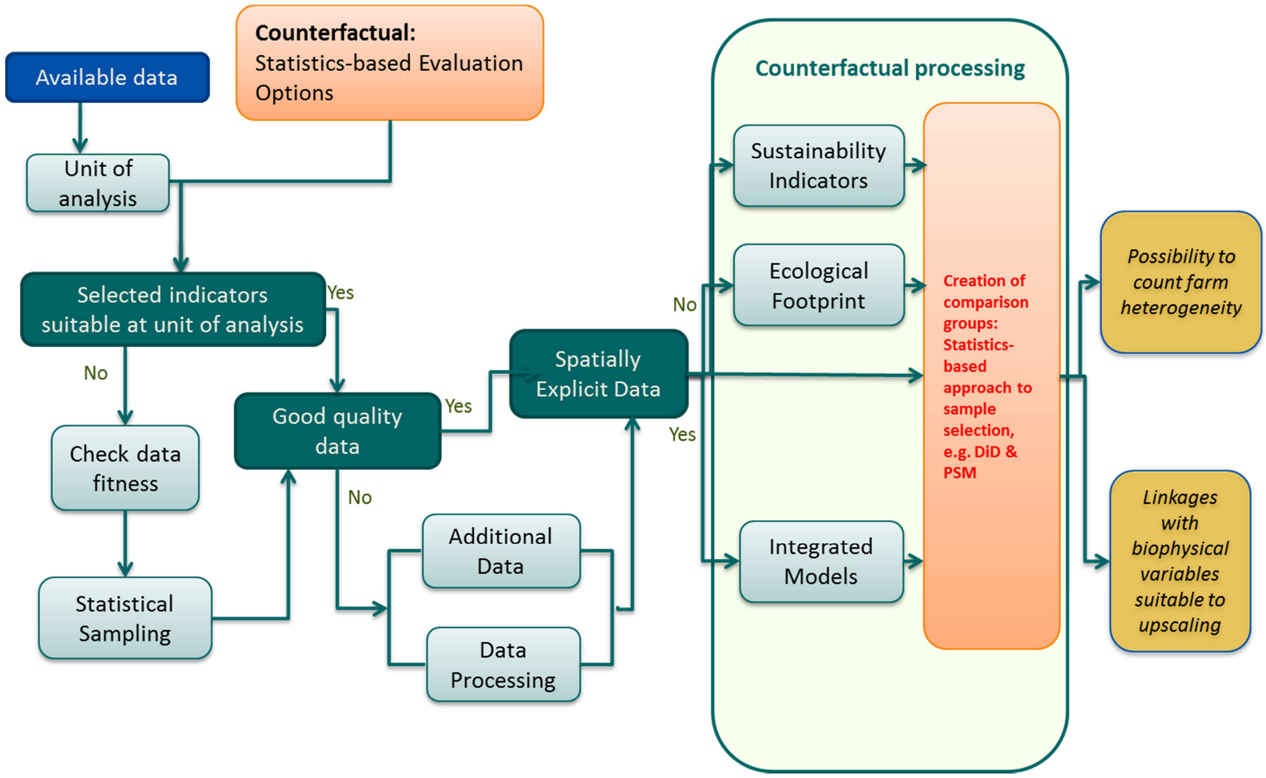

Figure 6 Elaborate Statistics-based Evaluation Options (Step 3.3c - micro level logic model)

3.3 Consistency with macro level

Step 3.4 - Micro-Macro aggregation and validation

In most of the cases the micro-macro level aggregation has to deal with multiple data sources, deriving from different databases with different metrics and terminology. Regarding the terminology, the farm can be defined as the baseline unit for micro-level analysis. However, it has to be underlined that ‘farm level’ can have different meanings in different evaluation exercises. An evaluation may use different scales, each with their own micro and macro level and therefore their use can be ambiguous. As highlighted before, in the evaluation assessment, micro level is substantially represented by the farm which is considered as the simplest management unit of the agricultural system linked to the implementation of RDP measures.

Each model can be more suitable for micro or macro-level evaluation if a consistent aggregation procedure is available for the analysis. From a micro-level point of view, spatial aggregation consists of up-scaling and aggregating data from farm level to regional or national levels. However, micro-macro linkage can be difficult to detect, in relation to the criticisms, in ensuring the representativeness of assessed data to the universe of farms. Although up-scaling could facilitate the consistency in micro-macro linkage aggregation, it has to be highlighted that the risk of summarising micro-level data to macro level cannot always be certain to represent the complexity of the universe of the agricultural systems.

Net impact evaluation at micro level can be ensured if indirect effects have been taken into account. In the case of environmental impacts at micro level, deadweight effects are relevant if land use and practice changes had occurred even without the intervention. Micro-macro linkages can lead to a better definition of indirect effects at macro level in presence of spatially explicit data in the case of environmental impact assessment.

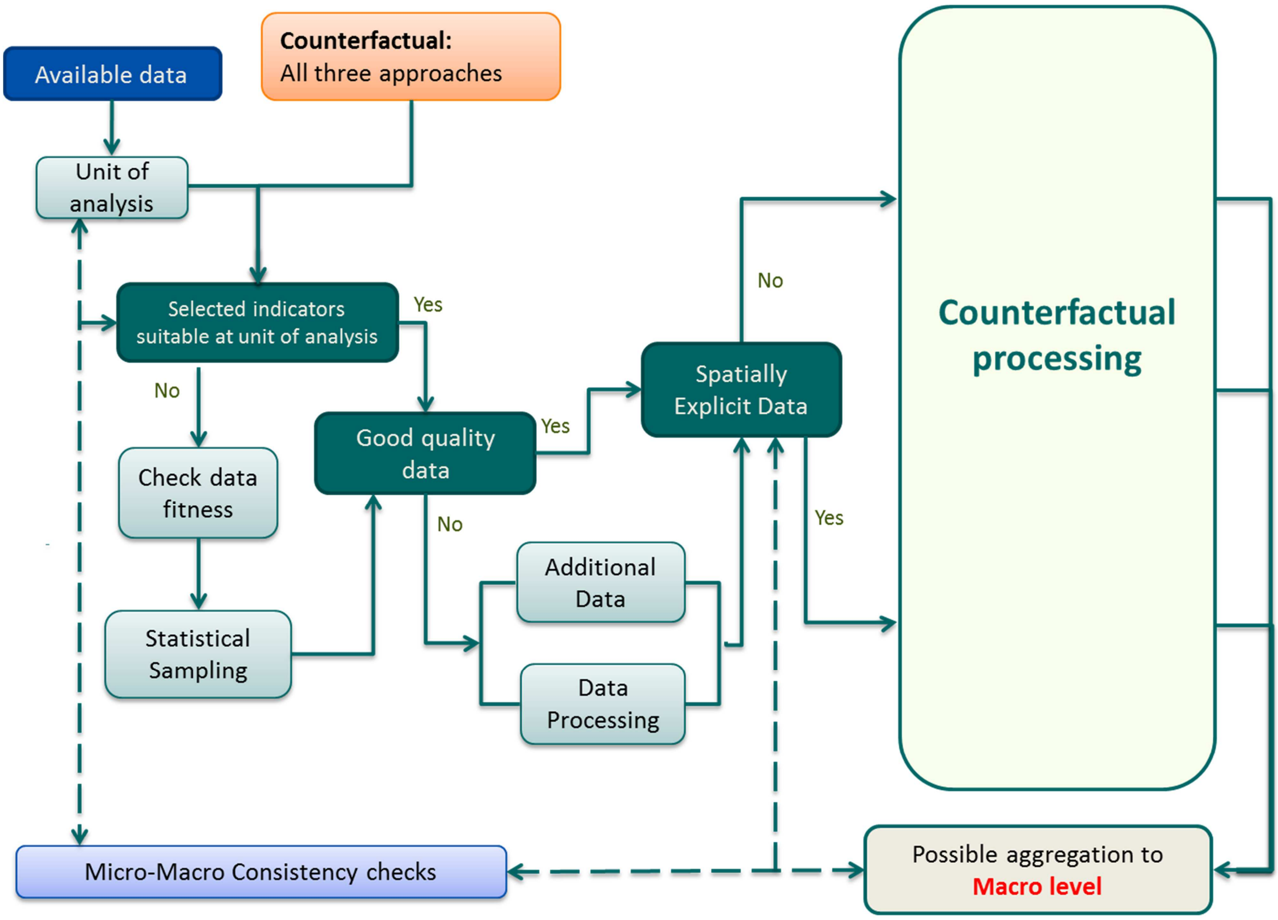

Figure 7 Micro-Macro aggregation and validation (Step 3.4 - micro level logic model)

4 Results from the case studies

4.1 Synthesis of the logic model development

The more relevant information about Data Requirement (DR) and Methodological Issues (MI) at micro level has been summarised below from the systematic list of issues included in the tables of Annex 1. The list has been structured per case study and options in terms of solutions and limitations.

Biodiversity High Nature Value Farmland, Italy

Indicator and Method tested: HNV Score as Composite indicator and Multicriteria Assessment (micro and macro level)

Solutions

- DR: At farm level the farming intensity can be easily estimated while for land cover the information on unfarmed features could be replaced by the extent of non-utilised agricultural areas and the presence of ecotones along small area of woods.

- MI: The FADN sample is the best sources of information; a geostatistical interpolation (Kriging method) has been used to define the probability maps on the regional distribution of HNV from farm level data.

Limitations

- DR: The available data are not sufficiently exhaustive, either in terms of the range of species covered, geographical coverage and ecological diversity, and they are not updated with sufficient regularity; for this reason additional surveys are necessary. Surveys at farm level for collecting data about semi-natural features could be very costly. Sampling design of FADN sample is not available at territorial level
- MI: Statistics-based models cannot be applied due to the low number of participant farms and the lack of detailed information on policy implementation, possibly overcome through a systematic link with the IACS database.

Biodiversity High Nature Value areas, Lithuania

Indicator and Method tested: Changes in diversity of ecotones and Spatial statistic to assess changes in landscape heterogeneity

Solutions

- DR: Quantitative impact of selected RDP measures was assessed. The selected sample sizes allowed measurement of effectiveness.
- To overcome the problem of temporal data, the combination of different data sources was utilised.
- MI: Not relevant

Limitations

- DR: The data used for the estimation of HNV does not allow to do the quality assessment of the impact.
- DR: Forest cadastre is renewed in different time periods across all country, that may cause some problems in terms of temporal dimension.
- MI: Not relevant

Biodiversity wildlife, Hungary

Indicator and Method tested: Farmland Bird Index and Difference-in-Difference method (micro and macro level)

Solutions

- DR: Based on the data available (bird census) it has been possible to select the indicator (farmland bird individuals at micro level). The sample size provided a good opportunity for the analyses of the selected case study without any problem connected with sample sizes.
- DR: For a robust biodiversity analyses assessment of longer time period is necessary to have access to time series data on Biodiversity 2009-2014 at micro level.
- MI: Control groups were selected based on the LPIS data. Environmental factors analysed have been based on CORINE Land Cover data.

- MI: Changes of FBI is proved to be a robust indicator of farmland biodiversity.

Limitations

- DR: Tests rarely involve the assessment of additional intervening factors (environmental, farmer behaviour), which may cause minor interpretation challenges.
- DR: For the designation of HNV a classification of natural areas based on farmer behaviour studies have been carried out.
- MI: Assessment of RD impacts based on the number of farmland bird individuals needs more detailed statistical analyses. Moreover, CLC data have limitations in a resolution necessary for micro level assessments.
- MI: CLC data have limitations in the resolution necessary for micro-level assessments.

Biodiversity wildlife, Lithuania

Indicator and Method tested: Corncrake (Crex crex) density and Multiple regression analysis

Solutions

- DR: Available data allow robust analysis of landscape stewardship scheme impacts. The approach collected data by public authority and no additional financial recourses are necessary. Adequate sample size provided the opportunity for a detailed landscape structure. Micro level allows quite detailed evaluation with samples of circular shape of 600 m diameter (0.28 km^2).
- MI: Available research data indicate a good robust dependency between the targeted measure and corncrake breeding timing.

Limitations

- DR: Data coverage is limited, available only within the projected areas boundaries, which are targeted for corncrake conservation. Data gathering does not consider land parcel structures, which would provide better conditions for the evaluation. Data availability limitation does not allow to perform macro level assessment
- MI: Timing of data gathering (second count) is too early to track impacts of late mowing restriction under the targeted measure.
- MI: Lacking statistical data on farmer behaviour.

Climate stability, Finland (only macro level)

Indicator and Method tested: GHG emission and General equilibrium model

Climate stability, Italy

Indicator and Method tested: GHG emission and Carbon Footprint (at process and farm level)

Solutions

- DR: The ‘Elaborate statistics-based’ analysis needs well-defined samples with a sufficient number of observations to perform regression analysis. The dimension of sample was sufficient at process level but not at farm level (both micro-level approaches).

DR: FADN sample (annually updated) and additional farm samples provide some data to estimate GHG emissions. Additional surveys have been preferred due to more specialised type of farming in the case study area that makes easier the use of the process level approach.

- MI: Carbon Footprint is a well-established method to estimate carbon emission from functional units having different structural and management characteristics.

Limitations

- DR: Only Qualitative and naïve quantitative analysis has been applied due to difficulties in data access (very long procedure to get access).
- DR: The choice of the process-level approach does not allow for time-based comparisons with the use of FADN.
- MI: Statistical significance of the parameters was not verified.

Landscape, Greece

Indicator and Method tested: Land cover change/visual amenity and Spatial analysis (land parcel and landscape level)

Solutions

- DR: For the case study, survey data are essential in order to ascertain the accuracy of remotely-sensed data. Spatial dimension was based on the classified land cover polygons (micro level).
- DR: The Google Earth images fit well with the temporal dimensions of the case study
- MI: Up-scaling of micro-level results was merely spatial.

Limitations

- DR: Lack of temporally differentiated participation data can increase probability of interpretation errors. IACS georeferenced data include information at land parcel level while classified land-cover polygons include more than one land parcels.
- DR: In terms of temporal dimension IACS georeferenced data are theoretically available every year. Google Earth images vary among area and time.
- MI: Land cover maps produced were not tested for their accuracy.
- MI: The unit of analysis was not linked to a programmatic scale.

Landscape, Scotland

Indicator and Method tested: Landscape Structural indicators and Landscape metrics (patch, class and landscape level)

Solutions

- DR: Land-cover monitoring data provide a detailed basis for the assessment of landscape structure indicators (baseline assessment). The data sample consists of spatial land-use data of the case study area, based on the IACS field boundaries. Land-use data recorded as part of IACS are available annually.

- MI: Land-use data (patch of land use) represents the smallest spatial unit for landscape metrics analysis. The micro-macro linkages are based on a coherent set of indicators for robust spatially-aggregated units (patch, class and landscape) that measure spatial structures in landscapes.

Limitations

- DR: IACS land use data have gaps in relation to non-agricultural land. IACS land use data is not accurate as land cover monitoring in relation to non-agricultural land use.
- MI: Limited resources meant that the assessment was limited to direct effects only.
- MI: The method can generate a large number of individual indicators.

Landscape, Scotland (only macro level)

Indicator and Method tested: Percentage of territory/UAA under Natura2000 and Spatial analysis

Soil quality, Hungary

Indicator and Method tested: Soil organic matter content and Sampling method

Solutions

- DR: One data source is used from the laboratory analyses.
- MI: More detailed sampling plan is needed and monitoring questions have to be formulated more precisely.

Limitations

- DR: Limitations are temporal at the time of writing in this specific case (soil organic matter content). Only a one-time sampling is available for monitoring purposes.
- MI: The large amount of data can be misleading as the higher the number the tighter the relationship between groups. The effect of soil organic carbon loss cannot be similar at flat and at steep slopes, so a huge number of samples from flat areas will show lower loss and little difference between 'with and without' farms and a high number of steep slopes will impact vice versa.

Soil quality, Scotland

Indicator and Method tested: Soil carbon and Biophysical modelling (by sub-catchment)

Solutions

- DR: RDP soil monitoring data can provide relevant data to support farm-level indicator data. Temporal land-use data can be derived from IACS land-use data which are annual and these are summarised to broad land-use classes.
- MI: The quantitative indicator does not explain the effects. Modelled changes for the indicator were used for DiD analysis.
- MI: Sub-catchment level can be considered as a micro level in relation to the European soil data which are available at NUTS3 level, i.e. the case study area. An aggregation of the results from sub-catchment to NUTS3 level can be validated against the EU level values of the indicator.

Limitations

- DR: The quality of the data for modelling is suitable only for analysis at sub-catchment level, and not for within sub-catchment (field/farm level) analysis. Currently soil monitoring takes place only at national level and not in a relevant temporal dimension for RDP assessment.
- DR: IACS land-use data are only available for agricultural land; hence there are gaps in the data for a range of other land uses (forestry, semi-natural areas and urban development). CORINE data are used to fill these gaps.
- MI: The IACS land use are not validated against Corine Land cover data or statistically tested for their accuracy.

Soil quality, Scotland

Indicator and Method tested: Soil erosion and Biophysical modelling (by sub-catchment)

Solutions

- DR: Data are available for the calculation of the indicator and creation of comparison groups. Temporal land use data can be derived from IACS land use data which are annual data and these are summarised to broad land use classes
- MI: The quantitative indicator does not explain the effects. Sub-catchment level can be considered as a micro level in relation to the European soil data which are available at NUTS3 level, i.e. the case study area. An aggregation of the results from sub-catchment to NUTS3 level can be validated against the EU level values of the indicator.

Limitations

- DR: The quality of the data for the modelling are suitable only for analysis at sub-catchment level, and not for within sub-catchment (field/farm level) analysis.
- DR: IACS land-use data are only available for agricultural land; hence there are gaps in the data for a range of other land uses (forestry, semi-natural areas and urban development). CORINE data are used to fill these gaps.
- MI: IACS land use data are not validated against Corine Land Cover data or statistically tested for their accuracy

Water quality - diffuse pollution, Finland (only macro level)

Indicator and Method tested: Nitrogen reduction (calculated GNB nitrogen) and Biophysical/Structural modelling

Water quality - diffuse pollution, Germany

Indicator and Method tested: Mineral N content in the soil in autumn (Nmin) and Pairwise comparison and Regression analysis (plot and water protection area level)

Solutions

- DR: Annual Monitoring data include a variety of variables on the farm structure and history and environmental conditions. Large samples of Nmin values were available at micro level which enabled the analysis to be conducted at sub-measure level. Data can be processed with common statistical software.

- MI: Nmin indicator is used in the analysis which is based on well-documented, theoretically-sound models and methods. Causal relationships have been quantitatively assessed through matching approach.

Limitations

- DR: Data was only provided as aggregated data set due to data protection reasons which can curb sound statistical analysis.
- DR: Micro level data is aggregated at the level of the drinking water protection area and cannot be used for the analysis.
- MI: Limited information on farm structure and management data did not allow the application of advanced matching techniques to further improve the analysis of causal relationships.
- MI: Statistical representativeness has not been verified. Only estimations of environmental impacts at macro level are possible.

Water quality - diffuse pollution, Germany

Indicator and Method tested: Gross nutrient balance (GNB) and Propensity Score Matching (farm level)

Solutions

- DR: Annual monitoring data on nutrient balances control data of the fertiliser ordinance, farm accountant data of the Land-Data Ltd, IACS data and farm structural data have been used.
- DR: Combinations of different data sources were explored to increase sample size.
- MI: Causal relationships have been quantitatively assessed through propensity score analysis. External assumptions have been implemented to improve consistency between results at micro and macro level.

Limitations

- DR: Combination of data from different sources is challenging because there isn't structural differences between the data sets. Some data sources (e.g. control data of the fertilizer ordinance) do only include net nitrogen balances.
- MI: not relevant for micro level.

Water quality - diffuse pollution, Greece

Indicator and Method tested: GNB and water use/ha and Biophysical model (land parcel and specific site of the NVZ of case study area)

Solutions

- DR: Use of existing data taking into account important crop types, soil conditions of the case study area in relation to the applied different farming practices of the AE action.

- DR: The biophysical model calculated the amount of nitrogen applied, as well as GNB in the form of nitrogen losses per ha. Moreover, it provided information on the irrigation rates that should be applied in order to avoid groundwater overexploitation.
- DR: The spatial dimension is based on the IACS field boundaries.
- MI: The biophysical model provided quantifiable results. Micro and macro linkages considered only in an intuitive manner.

Limitations

- DR: The Farm Accountancy Data Network data set lacks actual information on fertiliser application and/or water use. IACS georeferenced data are theoretically available every year on the contrary monitoring water quality and quantity data are irregular.
- MI: The obtained results were not verified with monitoring water quality and quantity data.
- MI: Farm level which is the decision level for participation in the various schemes was missing.

Animal Welfare, Germany

Indicator and Method tested: Integration of animal-based indicators (result-based elements) in a multi-criteria framework for the evaluation of animal welfare impacts (micro level)

Solutions

- DR: Secondary livestock and farm data are updated annually. Empirical monitoring data from farm visits were only available for one point in time. In case of long term evaluation contracts different sampling strategies can be explored to collect primary data through farm visits.
- MI: Conceptually and theoretically-sound models of the causal relationships could be developed for different relevant policy measures and animal welfare criteria and indicators.

Limitations

- DR: The available data are not sufficiently exhaustive to build large samples covering different livestock and farm types. Livestock monitoring through farm visits is very costly.
- MI: The quantitative assessment of the causal relationship depends on the availability of bigger samples of livestock monitoring data as well as access to existing livestock data.

4.2 Synthesis of the experiences (positive and negative)

The ‘Elaborate statistics-based’ assessment for HNV indicator needs well-defined samples with a sufficient number of observations to perform spatial analysis and regression models. The required data on semi-natural features, degree of farming intensity and presence of wild species linked to farmland are not sufficiently exhaustive, either in terms of the range of species covered, geographical coverage and ecological diversity, and they are not updated with sufficient regularity. When specific information is not available, proxy indicators are the alternative option. At farm level, the farming intensity can be easily estimated, while for land cover the information on unfarmed features could be replaced by the extent of non-utilised agricultural areas (HNV-IT) and the presence of ecotones along small area of woods (HNV-LT). For the analysis of HNV at micro level, additional collection of primary data is required mainly for semi-natural features and for the presence of wild species with a survey at two points in time. In this case the limitation is related to the cost of each single survey for collecting data on semi-natural features (HNV-IT).

The HNV indicator still lacks a well-recognised methodology for the estimation of the extent of HNV farmland. For the HNV evaluation, the temporal dimension constitutes an important factor and depends on the frequency of database updating. If databases are updated in different time periods, that may cause some problems in terms of time series analysis. To overcome this problem the combination of different data sources was utilised (HNV-LT). The micro-level sample about farmland birds provides a good opportunity for the analyses of Type 3 HNV farmland (BW-HU). The possible limitations are related to the classification of natural areas based on former studies.

The Climate Stability case studies are based on the GHG emission indicators estimated through the Carbon Footprint at process level and at farm level (CS-IT). Carbon Footprint is a well-established method to estimate carbon emissions from functional units having different structural and management characteristics. The data needed to estimate GHG emissions derived from existing farm surveys, such as FADN and FSS as well as additional surveys. The footprint approach requires specific data for the system referred to matter and energy flows. The complexity of the analysis increases with the complexity of the considered typology of production systems (e.g. mixed farms compared to mono-cultural farming systems). The process level survey has been preferred due to less demanding procedures for collecting and processing data and data collection is comparatively less expensive that the farm level one. Only qualitative and naïve quantitative analysis has been applied due to difficulties in data access (very long procedure to get access), while from the methodological point of view statistical significance of the parameters was not verified. Carbon Footprint widens the analysis of farming systems at fertilizer and energy sector and can be considered a more accurate indicator at farm level if compared to CMEF indicator.

Both case studies on landscape used the naïve DiD counterfactual with spatial analysis to assess land cover change/visual amenity (L-GR) and landscape structural indicator with landscape metrics (L-SCO). The method used in the Greek case study requires data between two observed periods (time series). The lack of temporally differentiated participation data and the accessed IACS data which did not include non-participant areas have further limited the assessment (L-GR). The groups were constructed from the remotely-sensed data (L-GR) or the data sample consisted of the spatial land-use data based on the IACS field boundaries (L-SCO). The land-cover change and visual amenity indicators include quantitative information but are not able to explain the effects as the analysis was limited only to the observed changes. On the other hand, the up-scaling of micro-level results was merely spatial (L-GR). In the landscape metrics approaches, the method has micro-macro linkages based on a coherent set of indicators for robust spatially-aggregated units (patch, class and landscape) that measure spatial structures in landscapes (L-SCO). The lack of temporally-differentiated participation data and the infrequency being out of sync with the RDP programme of land-cover monitoring are the main limitations in terms of data requirements. Other limitations are the lack of accuracy tests of the land-cover maps (L-GR) and the claims for measures commonly covering only part of a field (L-SCO).

The Soil Quality case studies are based on naïve quantitative as well as elaborate statistics-based evaluation approaches (SQ-SCO) to evaluate both soil quality and soil erosion indicators. Biophysical modelling was used for both indicators with the integration of spatial data on land use, topographic data and soil data. Due to limited data availability, the model was only suitable for analysis at sub-catchment level, and not for within sub-catchment (field/farm level) analysis (SQ-SCO). IACS land-use data, which are only available for agricultural land, presented gaps in the data for a range of other land uses (forestry, semi-natural areas and urban development). CORINE data are used to fill these gaps. Furthermore, RDP soil monitoring data can provide relevant data to support farm-level indicators; however, currently soil monitoring takes place only at national level and not in a relevant temporal dimension for RDP assessment (SQ-SCO). The method used for the soil carbon indicator aggregates the micro level (25m raster cell data) to sub-catchment level. In the case of the erosion indicator, the micro level (25m raster cell) is embedded in the method, because the location of a farm within a sub-catchment determines the actual impact on loss of soil through erosion (SQ-SCO). In the case of Hungary, elaborate statistics-based analysis has been performed, which needs a large soil sample size (SQ-HU). For the soil organic matter indicator, one data source is used, which includes data collected by the experts and data analysed in the laboratory. The method used shows the link between agricultural farming practices and soil quality. The lack of management data and the complexity of the system analysed (e.g. the effect of soil organic carbon loss cannot be similar in flat and steep slopes) require more detailed sampling plans. Aggregation of this micro-level data has to be statistically verified in terms of representativeness (SQ-HU).

The evaluation approach for Water Quality is based on the gross nitrogen balance (GNB) and mineral nitrogen indicator (Nmin). The methods can be used for both qualitative and naïve quantitative and elaborate statistics-based analysis. The data type could be derived from statistical sampling, requiring monitoring data and secondary data applied at the farm/parcel level, strictly dependent on sample size (FADN, CLC). The monitoring data is used that includes a variety of variables on the farm structure and environmental conditions for the Nmin indicator, while different data sources are used for the GNB approach (monitoring data, control data, LAND-Data Ltd, IACS data etc.). The Nmin indicator is used in the analysis and this is based on well-documented, theoretically-sound methods, while the GNB is a CMEF impact indicator and is used for the impact assessment of AEMs which is well-known and widely used for monitoring water quality. The Nmin indicator was selected as an additional impact indicator to evaluate the potential nitrate washed out into the groundwater. In terms of data requirements for Water Quality, the main limitation concerns the initial planned update of the analysis with micro-level data. However for recent years the micro-level data is aggregated at the level of the drinking water protection area and cannot be used for this kind of analysis.

For the estimation of GNB, the limited information on farm structure and management, along with lack of IACS data at different points in time, did not allow the application of advanced matching techniques to further improve the analysis of causal relationships (WQ-GR). Statistical representativeness has not been verified and only estimations of environmental impacts at macro level are possible. Both indicators, Nmin and GNB, should be used in combination for the same farms and sites in order to increase the validity of the analysis (WQ-DE). In general, it can be concluded that data availability and quality are the main limitations for the application of advanced methodological approaches as well as the construction of robust counterfactuals (WQ-DE).

The animal-based indicators can be used as part of a naïve quantitative assessment using ad-hoc methods to consider sample selection issues as well as through statistics-based analysis using explicit approaches for sample selection issues. For the statistics-based analysis, well-defined samples with a sufficient number of observations are needed for either secondary statistics or livestock monitoring data (farm visits) to perform robust matching methods such as propensity score matching at micro level. The analysis conducted through an integration of animal-based indicators in a multi-criteria assessment of animal welfare requires data on livestock health issues (e.g. lameness, mortality and body condition), housing and management conditions. The available data for the analysis are not sufficiently exhaustive to build large samples covering different livestock and farm types, and the livestock monitoring through farm visits is very costly. In the case of long-term evaluation contracts, different sampling strategies can be explored to collect primary data through farm visits. However, livestock monitoring data from 150 dairy farms covering different farm characteristics were available for the case study and additional empirical monitoring data from farm visits were only available for one point in time.

5 Conclusions

The conceptual framework presented in this report aims to systematise the *most commonly-used methodologies to assess micro-level environmental performance of agricultural policies. The rationale for this objective is based on the need to improve the overall objective of the evaluation, trying to select the appropriate indicators and methods and to deal with lack of monitoring and baseline data. An integrated approach - such as the logic model presented here - allows us to consider different perspectives of environmental issues from both quantitative and qualitative sides. It may help to valorise the use of existing methods and databases in order to fill the gap of knowledge on the agriculture-environment relationships within the complexity of multi-scale and multi-level approaches.

In a micro-level perspective it is important to consider the role of individuals and analyse in depth the different forms of organisation (spatial, networks, hierarchies) and interactions among different organisational and intervening levels. The choice of an appropriate scale has to represent both the natural processes and the socio-economic systems, in order to include multiple benefits and potential for cumulative environmental impacts. Only with multi-scale integration and the combination of results is it possible to efficiently generalise (up-scale) micro-level results in a macro-level perspective.

Field measurements, farm management surveys and farming system models essentially refer to the farm as the simplest management unit of an agricultural system, analysed from the point of view of a farmer who decides whether or not to participate in rural development schemes. This is one of the primary objectives of the logic model for the choice of the appropriate evaluation method, that is the measurement of net impacts of a policy measure at micro level when changes occurred even without the measure implementation (deadweight loss effects) or their behaviour induces changes in behaviour of other farmers in terms of practice changes (leverage effects). Looking at the experiences from case studies and other evaluation studies, the challenge is quite ambitious in terms of data availability for the creation of the control groups without selection bias (counterfactual approach) and identification of appropriate indicators that synthesise the causal links between farm inputs and outputs and environmental outcomes.

In most of the case study areas, almost only naïve quantitative analysis has been applied due to difficulties in data availability and data access, with negative effects from the methodological point of view when statistical significance of the parameters was not verified. The statistics-based approach to counterfactuals needs well-defined samples with a sufficient number of observations to perform regression models and spatial analysis.

The temporal dimension constitutes an important factor and depends on the frequency of database updating. The lack of temporally-differentiated participation data and the infrequency and being out of sync with RDP programme of the land cover monitoring are among the main limitations in terms of data requirements and it can further limit the assessment. If databases are updated in different time periods, that may cause some problems in terms of time series analysis.

Finally the lack of farm management data and the complexity of the system analysed requires a more detailed sampling plan. The limited information on farm structure and management data did not allow the application of advanced matching techniques to further improve the analysis of causal relationships. The data type could be derived from statistical sampling, requiring monitoring data and secondary data applied at the farm/parcel level, strictly dependent on sample size. However, when information from specific variables is not available, proxy indicators are the alternative option. For example, at farm level the farming intensity can be easily estimated while for land cover the information on unfarmed features could be replaced by some variables normally included in existing databases.

References

Aalders, I, Morrice, J, Miller, D, Schwarz, G (2015) *Report on the theoretical and methodological framework at macro level*. Report to the European Commission. Deliverable D5.3, ENVIEVAL project (Project Reference: 312071).

Artell, J, Aakkula, J, Toikkanen, H (2015) *Summary report on the methodological framework for counterfactual development.* Report to the European Commission. Deliverable D3.3, ENVIEVAL project (Project Reference: 312071).

De Benedetto L, Klemeš J (2009) The Environmental Performance Strategy Map: an integrated LCA approach to support the strategic decision-making process, *Journal of Cleaner Production* 17: 10.

Frankfort-Nachmias C, Nachmias D (1996) *Research Methods in the Social Sciences*, Fifth Edition, Edward Arnold, London.

Giupponi C, Carpani M. (2006) Recent developments in indicators and models for agri-environmental assessment. *Italian Journal of Agronomy* 1(4): 647-664.

Jax K. (2006) Ecological units: Definitions and application, *The Quarterly Review of Biology* 81(3): 237-258.

Millennium Ecosystem Assessment (2003) *Ecosystems and Human Well-Being: A Framework For Assessment*. Island Press, Washington.

Peters GP (2010) Carbon footprints and embodied carbon at multiple scales. *Current Opinion in Environmental Sustainability* 2: 245–250.

Povellato A, Forino G, Lasorella MV, Longhitano D (2013) *Review of micro level methods and scales*, Report to the European Commission. Deliverable D4.1, ENVIEVAL project (Project Reference: 312071)..

Primdahl J, Vesterager JP, Finn JA, Vlahos G, Kristensen L, Vejre H (2010) Current use of impact models for agri-environment schemes and potential for improvements of policy design and assessment. *Journal of Environmental Management* 91(6): 1245–1254.

Smith RD, Maltiby E (2001) *Using the Ecosystem Approach to Implement the CBD (Global Synthesis Report)*. University of London, London.

Van der Werf HM, Petit J (2002) Evaluation of the environmental impact of agriculture at the farm level: a comparison and analysis of 12 indicator-based methods. *Agriculture, Ecosystems and Environment* 93(1-3): 131–145.

Weber J-L (2011) An Experimental Framework for Ecosystem Capital Accounting in Europe, *EEA Technical report* No 13/2011. European Environment Agency, Copenhagen,

Weber J-L (2014) Ecosystem Natural Capital Accounts: A Quick Start Package, *Technical Series* No. 77, Secretariat of the Convention on Biological Diversity, Montreal.

Weidema B, Wenzel H, Petersen C (2004) The Product, Functional Unit and Reference Flows in LCA. *Environmental News* No. 70, Danish Environment Protection Agency, Copenhagen.

Whittaker C, McManus MC, Smith P (2013) A comparison of carbon accounting tools for arable crops in the United Kingdom. *Environmental Modelling & Software* 46: 228-239.

Annex 1

Public good: Biodiversity High Nature Value Farmland, Italy (HNV-IT)
Indicator and Method tested: HNV Score as Composite indicator and Multicriteria Assessment (micro and macro level)

Dimensions	**Pre-conditions**	**Solutions**	**Limitations**
Integration of counterfactuals	The Biodiversity HNV method can be used for both "Qualitative and Naive quantitative" and "Elaborate statistics-based" analysis	The "Elaborate statistics-based" analysis needs well defined samples with sufficient number of observations to perform spatial analysis and regression models to aggregation and scaling up of micro level (farm).	The dimension of FADN sample was only sufficient for Naive quantitative analysis while the analysis at macro level has been only carried out for one year.
DATA REQUIREMENT ISSUES			
Type of data	The HNV indicator requires specific data on land cover, intensity of the farming systems and presence of wild species linked to farmland	Proxy indicators are the alternative option when specific information is not available. - At farm level the farming intensity can be easily estimated while for land cover the information on unfarmed features could be replaced by the extent of non-utilised agricultural areas and the presence of ecotones along small area of woods. - At regional level IACS-LPIS represent an important source of information for land cover, although n	The available data are not sufficiently exhaustive, either in terms of the range of species covered, geographical coverage and ecological diversity, and they are not updated with sufficient regularity.
Primary monitoring data	Additional collection of primary data is required mainly for semi-natural features and for the presence of wild species with a survey in two times	Not easy to find proxy indicators with good information value	Surveys at farm level for collecting data about semi-natural features are very costly.
Sample size	RDP participants in FADN samples at regional level are not sufficiently representative for aggregation to macro level and use of elaborate statistics-base models		
Spatial dimension	HNV can be better estimated with spatially	The spatial distribution of the FADN sample	Presence of unfarmed features (mainly the

	explicit data	can be monitored through IACS-LPIS database	linear ones) is hardly detected in the available GIS database (such as Corine Land Cover)
Temporal dimension	The analysis needs, at least, data from two different years at the beginning and at the end of the RDP programming period	FADN sample is annually updated	GIS databases generally maintained by monitoring agencies are updated with no reference to RDP programming phases
Processing requirements			
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	Need to have well-documented, theoretically sound models and methods that show the links between farming practices and environmental impacts	The HNV indicator still lacks of a well-recognised methodology for the estimation of the extent of HNV farmland. The replication of methodology adopted in other case studies allows for a first measurement of RDP impact on HNV farmland	
Assessment of net-impact	The estimation of direct and indirect effects needs the availability of control groups	- The FADN sample is the best sources of information - IACS-LPIS data are potentially available with yearly update	- The FADN sample does not have detailed information on policy implementation and the number of participant farms is too low - IACS-LPIS databases are not implemented for an easy use for statistical analysis
Micro-macro linkage	The aggregation of micro level data has to be statistically verified in terms of spatial representativeness. The data sources used for micro-macro level evaluation, using different metrics and terminology, have to be adapted before upscaling from farm to regional level. Spatially aggregated units (patch, class and landscape) that measure spatial structures in landscapes are needed for a net impact assessment.	The aggregation of FADN data at regional level has been realised using geostatistical interpolation (Kriging method) to define the probability maps on the regional distribution of HNV from farm level data	
ADDITIONAL COMMENTS			

Public good: Biodiversity High Nature Value areas, Lithuania (HNV-LT)
Indicator and Method tested: Changes in diversity of ecotones and Spatial statistic to assess changes in landscape heterogeneity

Dimensions	**Pre-conditions**	**Solutions**	**Limitations**
Integration of counterfactuals	The spatial statistic is applicable for "Qualitative and Naive quantitative" analysis and before and after counterfactuals was selected.	Good quality spatial data needed to be able to capture impact before and after.	The approach has limitations for quality assessment
DATA REQUIREMENT ISSUES			
Type of data	Spatial and statistical IACS data, Forest cadastre data, Georeferential spatial data set at scale 1:10 000 of the Republic of Lithuania (GDR10LT), Ortophoto maps.	Quantitative impact of selected RDP measures was assessed	This data does not allow to do the quality assessment of the impact.
Primary monitoring data	Such data was not used	-	-
Sample size	Parcel size of the declared plot in micro level and geographical region in macro level	The selected sample sizes allowed to make measure effectiveness calculations	In the macro level, measure effectiveness showed the total effectiveness of the measure, but you couldn't estimate certain parcels.
Spatial dimension	Indicators are estimated with spatially explicit data.	Spatial dimension was based on IACS data parcels.	-
Temporal dimension	Temporal dimension depends on the frequency of data bases	Temporal dimension problem was eliminated combining different data bases	Forest cadastre is renewed in different time periods across all country, that may cause some problems in the future. GDR10LT database is based on the Ortophoto images, so it is produce a year later than the images themselves. However having good IACS data decreases the degree of these problems
Processing requirements	Processing requires GIS software	Data was processed using standard GIS	
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	The method and spatial IACS data allow to establish robust causal relationship on both micro and macro levels.	Not relevant	

Assessment of net-impact	The method allows to exclude an impact of non relevant measures or other policies impacts.	Not relevant	
Micro-macro linkage	Good resolution spatial data on micro level do allow upscaling to macro level	Not relevant	
ADDITIONAL COMMENTS			

Public good: Biodiversity wildlife, Hungary (BW-HU)

Indicator and Method tested: Farmland Bird Index and Difference-in-Difference method (micro and macro level)

Dimensions	**Pre-conditions**	**Solutions**	**Limitations**
Integration of counterfactuals	Based on the data available (bird census) the selected indicators (FBI-macro, farmland bird individuals) provide the possibility of using combined before-after and with-without comparisons.	Parallel spatial analyses of the bird census data and the RD (214- agri-environment measures) uptake data gave enough detailed samples for carry out the test of the selected method.	Tests rarely involve the assessment of additional intervening factors (environmental, farmer behaviour), which may cause minor interpretation challenges.
DATA REQUIREMENT ISSUES			
Type of data	Biodiversity: Common Birds Monitoring Program carried out by BirdLife Hungary RD uptake data: land parcels under AE contracts from LPIS Land cover data: based on widely available data sources (eg. CORINE)	Overall examination of the available data provided good opportunity for the analyses.	Forming participant and non-participant groups faced with challenges, as spatial selection of biodiversity survey spots have not followed the spatial distribution of RD AE contracted parcels.
Primary monitoring data	Additional data collection was not necessary	Not relevant	Not relevant
Sample size	Macro level samples were representative at country level, while micro level sample size provided good opportunity for the analyses of the selected case study area	Sample sizes have not faced with problems	Designing participant-non-participant groups needed detailed analyses.
Spatial dimension	Macro level: country Micro level: Heves-plain case study area	Biodiversity data is collected in 300-400 survey squares/year LPIS data was available for 2009-2014 time period Land use data have been assessed based on Corine Land Cover.	Classification of natural areas based on former studies aiming High Nature Value Area designation in Hungary
Temporal dimension	For a robust biodiversity analyses assessment of a longer time period is necessary	Biodiversity data was available for 1999-2014 at macro level, while 2009-2014 in micro level	-
Processing requirements	GIS analyses is necessary	GIS analyses were carried out aiming the definition of participant-non participant groups (at both level) and natural-not natural survey spots (micro level)	-

METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	Need to have well-documented, theoretically sound models and methods that show the links between farming practices and environmental impacts	Changes of FBI is proved to be a robust indicator of farmland biodiversity	Assessment of RD impacts based on the number of farmland bird individuals needs more detailed statistical analyses.
Assessment of net-impact	The estimation of direct and indirect effects needs the availability of control groups	- In terms of participation in AE measures control groups selected based on the LPIS data, while environmental factors analysed based on CORINE Land Cover data	CLC data have limitations in a resolution necessary for micro level assessments
Micro-macro linkage	The aggregation of micro level data has to be statistically verified in terms of spatial representativeness.	Not relevant.	
ADDITIONAL COMMENTS			

Biodiversity wildlife, Lithuania

Indicator and Method tested: Corncrake (Crex crex) density & Multiple regression analysis

Dimensions	Pre-conditions	Solutions	Limitations
Integration of counterfactuals	The spatial statistic applicable for the analysis, available data allow to design with-without comparisons.	Parcel spatial analysis of the corncrake monitoring data combined with AEM (214-agri-environment measures) uptake data gave enough detailed samples for carry out the test of the selected method.	Tests rarely involve the assessment of additional intervening factors (environmental, farmer behaviour), which may cause minor interpretation challenges.
DATA REQUIREMENT ISSUES			
Type of data	- Biodiversity: Corncrake singing males monitoring data - RD uptake data: land parcels under AE contracts from LPIS	Available data allows is responsive for robust analysis of landscape stewardship scheme impacts	Data coverage is limited, available only within the projected areas boundaries, which are targeted for corncrake conservation. Data gathering does not consider land parcel structures, which would provide better conditions for the evaluation.
Primary monitoring data	Corncrake singing males density data was taken by state biodiversity monitoring program.	The approach allows to use already gathered data by public authority, no need of substantial additional financial recourses for data gathering; Data source is constancy updated as part of state monitoring program.	The data is primarily used for biodiversity status evaluation and not for RDP assessment. Therefore, spatial coverage does not correspond to RDP assessment needs. Some data gathering aspects needs improvement (.g. timing of monitoring, land parcel structure record).
Sample size	Micro level samples were formed within the case study area. Sample size 0,28 km2	Sample size, provided opportunity for detailed look and smooth landscape structure	Sample shape was determined by the corncrake data gathering observation point (circular shape) and in some cases did not allow to include full land parcels.
Spatial dimension	Indicators are estimated with spatially explicit data.	Micro level allows quite detailed evaluation with samples of circular shape of 600 m diameter (0,28 km2)	Data availability limitations does not allow to perform macro level assessment
Temporal dimension	Temporal dimension covering long period with annual frequency of data	Evaluation covered only 20107 data	Sort temporal dimension does not allow to form before-after comparison groups and does not trace effects from previous seasons of the

			breeding success and migration impact on birds density
Processing requirements	GIS analyses is necessary	Data was process using standard features of GIS	-
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	Available and well documented research results on the breeding success depending on the timing of mowing. Timing of data gathering is close to the timing of mowing restrictions.	Available research data indicates a good robust dependency between the targeted measure and corncrake breeding timing	Timing of data gathering (second count) is too early to track impacts of late mowing restriction under the targeted measure
Assessment of net-impact	The estimation indirect effects needs data availability of farmers behaviour and other indirect environmental factors (e.g. migration mortality rate)	Multiple regression analysis model allows to consider direct environmental impacts	Lacking statistical data on farmer behaviour.
Micro-macro linkage	Was not tested	Not relevant.	
ADDITIONAL COMMENTS			

Public good: Climate stability, Finland (CC-FI)

Indicator and Method tested: GHG emission and General equilibrium model

Dimensions	Pre-conditions	Solutions	Limitations
Integration of counterfactuals	The DREMFIA sectoral economic model allows a number of counterfactuals. The evaluator needs to decide which type of counterfactuals are required.	Construct a number of counterfactuals that could have been politically viable in the absence of agri-environmental payments.	
DATA REQUIREMENT ISSUES			
Type of data	The DREMFIA sectoral economic model uses a wide variety of data (input and output prices, demand for agricultural produce among others) that are continuously collected and updated into the model.	Upkeep of data sources to be fed into the model.	Sudden changes in data availability may be problematic for model use.
Primary monitoring data	Environmental effects (CO_2-equivalent measure) are calculated within the model using transfer functions. Model produces also information on other environmental indicators (e.g. fertilizer use).		
Sample size	n/a, the model calculates a regions as representative farms		
Spatial dimension	Data used is national averages, while analysis is conducted on a regional level		The model is strictly macro-level analysis.
Temporal dimension	Annual		The choice of the process level approach does not allow for time-based comparisons,
Processing requirements	A sectoral economic model requires experts building, upkeeping and being able to conduct analysis using the model.	Upkeeping the staff and data to keep the model usable.	Risks of significant changes in personnel and data availability.
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	The model builds on profit maximizing farmer behaviour and environmental impacts through		

	farming pressure are based on transfer functions identified in the literature.		
Assessment of net-impact	Decision of relevant counterfactuals by the evaluator.	Choosing politically viable alternative(s) of counterfactual(s).	Macro-level model does not identify local impacts, though in the case of a global pollutant, the issue is less severe.
Micro-macro linkage	The DREMFIA model is strictly a macro-level model that builds on an aggregate micro-level farm response.		
ADDITIONAL COMMENTS			

Public good: Climate stability, Italy (CC-IT)

Indicator and Method tested: GHG emission and Carbon Footprint (at process and farm level)

Dimensions	**Pre-conditions**	**Solutions**	**Limitations**
Integration of counterfactuals	The Carbon Footprint method can be used for both "Qualitative and Naive quantitative" and "Elaborate statistics-based" analysis	The "Elaborate statistics-based" analysis needs well defined samples with sufficient number of observations to perform regression analysis. The dimension of sample was sufficient at process level but not at farm level	Only Qualitative and Naive quantitative analysis has been applied due to difficulties in data access (very long procedure to get access)
DATA REQUIREMENT ISSUES			
Type of data	The footprint approach requires specific data for the system referred to matter and energy flows. The complexity of the analysis increases with the complexity of the considered typology of production systems (e.g. mixed farms compared to mono-cultural farming systems)	- Farm surveys, such as FADN/FSS, already provides some data needed to estimate GHG emissions. - The process level surveys has been preferred due to less demanding procedures for collecting and processing data - The high frequency of specialised type of farming in the case study area makes easier the use of the process level approach	
Primary monitoring data	Additional collection of primary data is required on: - Input use and yields - Information of farm practices - Structural data	Process-level data collection is comparatively less expensive that farm level one	
Sample size	Samples of participants and non-participants process/farms sufficiently representative for aggregation to macro level and use of elaborate statistics-base models	The need of a farm sample larger than the already available FADN, due to low representativeness of the participant group and the high differentiation of the structural characteristics, have leaded to opt for the process level approach, less demanding in terms of units of observations	
Spatial dimension	GHG emissions can be better estimated with spatially explicit data	The spatial distribution of the sample has been designed with reference to the distribution of participants based on IACS/LPIS maps	

Temporal dimension	Input-output tables periodically updated	FADN sample is annually updated	The choice of the process level approach does not allow for time-based comparisons,
Processing requirements	CF can be considered as a subset of data derived from LCA approach. LCA is based on International Standards (ISO 14040, ISO 14044) and on environmental labels and declarations (ISO 14020, ISO 14024, ISO 14025)	Data can be processed with common statistical software	
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	Need to have well-documented, theoretically sound models and methods that show the links between farming practices and environmental outcomes	Carbon Footprint is a well-established method to estimate carbon emission from functional units having different structural and management characteristics	
Assessment of net-impact	The estimation of direct and indirect effects needs the availability of control groups	Sample size of treated and non-treated groups has been adjusted to ensure robust estimation	Statistical significance of the parameters was not verified
Micro-macro linkage	The aggregation of micro level data has to be statistically verified in terms of representativeness	The aggregation at regional level has been realised using coefficients to include all the crops not analysed at micro level.	Statistical representativeness has not been verified
ADDITIONAL COMMENTS			

Public good: Landscape, Greece (L-GR)

Indicator and Method tested: Land cover change/visual amenity and Spatial analysis with geo-statistical approach (land parcel and landscape level)

Dimensions	Pre-conditions	Solutions	Limitations
Integration of counterfactuals	The Spatial analysis with geostatistical approach can be used for "Elaborate statistics-based evaluation options" analysis.	The DiD analysis compares the before and after changes of programme between participants and non participants. Method requires data availability between two periods observed (time series).	Partial information on participants and non-participants. DiD analysis was limited only to the changes observed. Small number of observations among the comparison groups.
DATA REQUIREMENT ISSUES			
Type of data	IACS georeferenced data, remote sensed data drawn from GE images.	Non-participant group was only constructed from the remote sensed data.	Lack of temporally differentiated participation data. Accessed IACS data did not include non participant areas.
Primary monitoring data	Land cover data	Ground truth survey data are essential in order to ascertain the accuracy of remote sensed data.	Increased probability of interpretation errors.
Sample size	Large	Spatial analysis does not use farm samples but all land parcels that are supported by the examined measures (inventory).	
Spatial dimension	Indicators are estimated with spatially explicit data.	Spatial dimension was based on the classified land cover polygons.	Although IACS georeferenced data include information at land parcel level, the classified land cover polygons consist of more than one land parcels, since the manual digitisation was processed according to the neighbouring features.
Temporal dimension	The temporal dimension strictly depends on the frequency of land cover data.	Dates of capture of GE images fit well with the temporal dimensions of our examined measure.	IACS georeferenced data are theoretically available every year. GE images vary among area and time.
Processing requirements	Processing requires spatial analytical /GIS skills.	Data are being processed using standard GIS software.	
METHODOLOGICAL			

ISSUES			
Establishment of robust causal relationships	Need to have well-documented, theoretically sound models and methods that show the links between farming practices and environmental outcomes	Indicators include quantitative information but are not able to explain the effects.	Land cover maps produced were not tested for their accuracy. Neither statistical test nor regression analysis was conducted
Assessment of net-impact	The estimation of direct and indirect effects needs the availability of control groups	DiD analysis was limited only to the observed changes.	Farm level which is the decision for participation in the various schemes was missing. The functional unit was not linked to a programmatic scale.
Micro-macro linkage	Macro level can build on micro level analysis	Up-scaling of micro level results was merely spatial.	Farm level which is the decision for participation in the various schemes was missing. The functional unit was not linked to a programmatic scale.
ADDITIONAL COMMENTS			

Public good: Landscape, Scotland (L-SCO)

Indicator and Method tested: Landscape Structural indicators and Landscape metrics (patch, class and landscape level)

Dimensions	**Pre-conditions**	**Solutions**	**Limitations**
Integration of counterfactuals	The Landscape Metrics can be constructed for a "Elaborate statistics based" analysis	The DiD analysis compares the before and after changes of programme between participants and non-participants. Method requires data availability between two periods observed (time series). If the trend of the data are the same it will allow an elaborate statistics based assessment, otherwise it will be a naïve quantitative comparison.	There are before and after data for participants and non-participants however the data at this stage of the testing of the indicators do not have the quality required for elaborate statistical approach. Therefore the assessment if using a naïve qualitative comparison.
DATA REQUIREMENT ISSUES			
Type of data	The landscape metrics require spatial data regarding land use or land cover. The method is case sensitive so the resolution of the data in the comparison need to be based on land use/land cover data with the same data quality.	Land cover monitoring data provide a detailed basis for the assessment of landscape structure indicators (baseline assessment). Land use data recorded as part of IACS data base are an alternative source of information	Land cover monitoring is infrequent and out of synch with RDP programme IACS land use data have gaps in relation to non-agricultural land
Primary monitoring data	No primary data used		
Sample size	Large	The data sample consist of the spatial land use data of the case study area, based on the IACS field boundaries.	Reporting of IACS land use data is not as accurate as land cover monitoring in relation to non-agricultural land use which does have an impact on the landscape structure. Land cover data (CORINE) would be more suitable
Spatial dimension	The method can be applied to different spatial levels. Given the data dependency of the method ideally the minimum mappable area of the indicator should determine the level.	The spatial dimension is based on the IACS field boundaries.	
Temporal dimension	Monitoring frequency approximately per decade but not in sync with RDP programme cycle,	By using IACS land use data rather than land cover monitoring data the data are available annually Creation of land cover maps using remote	The land use data can only be summarised by field boundary This process is costly and demanding and will

		sensing data	require field validation before use and may only be feasible for relatively small geographic areas.
Processing requirements	Processing requires spatial analytical /GIS skills. Creation of time series by data updating using RS.	Data are being processed using standard GIS software	
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	The robust causal relationship requires land use/cover change due to measures to be recorded.	Land use data recorded by field, which are used as the basis of a patch of land use, which is the smallest spatial unit for landscape metrics analysis.	Claims for measures commonly cover only part of a field, which means that there is an over-estimation of the area under measure.
Assessment of net-impact	The estimation of direct and indirect effects needs the availability of control groups	The data made it possible to conduct a DiD analysis to assess change in the indicators for the comparison groups (before/after and with/without participation)	The DiD analysis is using the mean of the indicator value for each comparison group, but the data do not have enough detail to explain the participation and non-participation.
Micro-macro linkage	Macro level can build on micro level analysis	The method has a micro-macro linkages based on a coherent set of indicators for robust spatially aggregated units (patch, class and landscape) that measure spatial structures in landscapes.	The method can generate a large number of individual indicators, which need to tested to assess their suitability for use in RDP impact assessment.
ADDITIONAL COMMENTS			

Public good: Landscape, Scotland (L-SCO)

Indicator and Method tested: Percentage of territory/UAA under Natura2000 and Spatial analysis

Dimensions	Pre-conditions	Solutions	Limitations
Integration of counterfactuals	For the indicator Natura2000 only a naïve comparison can be constructed	Natura2000 as measure of impact of RDP on landscape has limitations both as a measure and for the creation of counterfactual design. Alternative means of measuring Natura2000 in relation to agricultural land could be considered.	In the case study area Natura2000 is static hence assessing the impact of RDP on landscape based on Natura2000 is limited.
DATA REQUIREMENT ISSUES			
Type of data	The data are Natura2000 boundary data, UAA areas and IACS measure uptake data.	The available data allowed an assessment of change in RDP measures supporting Natura2000 areas.	These data do not really provide information for the impact assessment on landscape.
Primary monitoring data	No primary data used		
Sample size	All Natura2000 areas	The data sample consist of the spatial data of Natura2000 in the case study area and the UAA areas based on the IACS data.	Natura2000 are static for the RDP period and the value of the indicator changes due to changes in UAA area and RDP measure uptake
Spatial dimension	UAA and area of territory	The spatial dimension is based on the Natura2000 and UAA boundaries.	Calculated change does not measure change in Natura2000, but change in the amount of Natura2000 classified as UAA.
Temporal dimension	Natura2000 does not change much and for the RDP programme cycle it is static	The area of UAA does undergo minor changes from year to year. The main change for the case study is in the uptake of RDP measures	Calculated change does not measure change in Natura2000, but change in the amount of Natura2000 classified as UAA for participants and non-participants
Processing requirements	Processing requires basic GIS	Data are being processed using standard GIS software	
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	It is not possible to establish a robust causal relationship, for this proposed CMES indicator for landscape.	Consider alternative measure of Natura2000 for the relationship between RDP and landscape	Given that areas under Natura2000 are not changing much this does not seem to be a good indicator for this approach.

Assessment of net-impact	The estimation of direct and indirect effects needs the availability of control groups	Based on the data it is possible to construct a DiD analysis of change in the relationship between UAA with Natura2000 and the extent RDP is contributing the maintenance.	The data do not explain the participation and non-participation, nor is it a real measure of change.
Micro-macro linkage	The nature of the indicator is a macro level indicator only	An alternative indicator needs to be considered	This will require a range of different possible alternatives for measuring the impact of RDP on landscape through Natura2000. Unfortunately this was beyond the scope of this project.
ADDITIONAL COMMENTS			

Soil Quality, Hungary (SQ-HU)*

Indicator and Method tested: Soil organic matter content and Sampling method

Dimensions	Pre-conditions	Solutions	Limitations
Integration of counterfactuals	Large soil sample size can be used for "Elaborate statistics-based evaluation options" analysis of soil organic matter.	The "Elaborate statistics-based" analysis needs large number of samples at national scale for macro level to perform statistical analysis.	Only “with and without” analyses could be applied as there was no data from the farms before they started the programs.
DATA REQUIREMENT ISSUES			
Type of data	Samples collected by experts and analysed in qualified laboratories.	One data source is used, the data from the laboratory analyses.	The non.-participating farms lack certain environmental data that is collected on participating farms.
Primary monitoring data	Plot level data.	One time sampling of soil information on “with and without farms” have been used.	Limitations are temporal in the time of writing in this specific case (soil organic matter content).
Sample size	Large.	NA	NA
Spatial dimension	National.	Data need to be collected from non-participating farms.	Data is missing from non-participating farms.
Temporal dimension	Only a one-time sampling is available for monitoring purposes.	Before and after sampling is needed to conduct Difference-in-Difference method that provides better results and performance.	Only a one-time sampling is available for monitoring purposes.
Processing requirements	Soil data is examined with statistical analyses.	NA	NA
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	It is based on methods to show the link between agricultural farming practices related to soil information.	Data need to be collected from these farms as well.	The lack of management data from non.-participating farms limited the analyses.
Assessment of net-impact	The estimation of net impacts is based on control groups.	There has to be a before and after sampling.	Temporal sampling constraints did not allow to apply a DiD counterfactual approach to find changes between different comparison groups over time.
Micro-macro linkage	The aggregation of micro level data has to be	More detailed sampling plan is needed and	The large number of data can be misleading as

	statistically verified in terms of representativeness.	monitoring questions have to be formulated more precisely.	the higher the number the tighter the relation between groups. On the other hand, as soil has many properties, there were certain missing comparison groups for certain soil properties for a better evaluation. E.g. the effect of soil organic carbon loss cannot be similar at flat and at steep slope, so a huge number of samples from flat areas will prove lower loss and little difference between “with and without” farms as high number of steep slopes will impact vice versa.
ADDITIONAL COMMENTS	The pre-conditions is basically policy (and money) driven. More scientific input might be needed and as there is normally lack of time, more time would be needed for preparation of monitoring activities.	The solutions are basically originated from the limitations this is why a list of limitations must be well-analysed and also, included in the development of the logic model.	The limitations are normally policy-driven as well. Again, more scientifically based approach is needed for the planning of the monitoring.

Public good: Soil quality, Scotland (SQ-SCO)

Indicator and Method tested: Soil carbon and Biophysical modelling (by subcatchment)

Dimensions	**Pre-conditions**	**Solutions**	**Limitations**
Integration of counterfactuals	Naïve quantitative as well as elaborate statistics-based evaluation approaches can be used	The DiD analysis compares the before and after changes of programme between participants and non-participants. Method requires data availability between two periods observed (time series). If the trend of the data are the same it will allow an elaborate statistics based assessment, otherwise it will be a naïve quantitative comparison.	While there are before and after data for participating and non-participating sub-catchments the data do not have the quality that will allow for more than a naïve qualitative comparison.
DATA REQUIREMENT ISSUES			
Type of data	Spatial data land use, topographic data, and soil data	Data are available for the model the indicator and subsequent creation of comparison groups, based on sub-catchment with and without participation.	The quality of the data for the modelling are suitable only for analysis at sub-catchment level, and not for within sub-catchment (field/farm level) analysis.
Primary monitoring data	No primary data used		
Sample size			
Spatial dimension	Indicator is calculated by sub-catchment	RDP relevant soil monitoring data can provide relevant data to support farm level indicator data.	Currently soil monitoring takes place only at national level and not in relevant temporal dimension for RDP assessment.
Temporal dimension	The availability of temporal data is dependent on the source of the data, Corine land cover data are not in sync with RDP programme cycle	Temporal land use data can be derived from IACS land use data which are annual data and these are summarised to broad land use classes.	IACS land use data are only available for agricultural land hence there are gaps in the data for a range of other land uses (forestry, semi-natural areas and urban development). CORINE data are used to fill these gaps.
Processing requirements	GIS software	Data processed in standard GIS using scripts	
METHODOLOGICAL ISSUES			
Establishment of robust	Method used is based on robust and well	The quantitative indicator does not explain	The IACS land use are not validated against

causal relationships	documented relationship between land use and soil carbon for the area.	the effects	Corine Land cover data or statistically tested for their accuracy
Assessment of net-impact	The estimation of direct and indirect effects needs the availability of control groups	Modelled changes for the indicator were used for DiD analysis.	The DiD is using the mean of the indicator value by sub-catchment and is unable to explain the participation and non-participation.
Micro-macro linkage	The method aggregates the micro level (25m raster cell data) to sub-catchment level.	Sub-catchment level can be considered as a micro level in relation to the European soil data which are available at NUTS3 level, i.e. the case study area. An aggregation of the results from sub-catchment to NUTS3 level can be validated against the EU level values of the indicator.	
ADDITIONAL COMMENTS			

Public good: Soil quality - Scotland (SQ-SCO)

Indicator and Method tested: Soil erosion and Biophysical modelling (by sub-catchment)

Dimensions	Pre-conditions	Solutions	Limitations
Integration of counterfactuals	Naïve quantitative as well as elaborate statistics-based evaluation approaches can be used	The DiD analysis compares the before and after changes of programme between participants and non-participants. Method requires data availability between two periods observed (time series). If the trend of the data are the same it will allow an elaborate statistics based assessment, otherwise it will be a naïve quantitative comparison.	While there are before and after data for participating and non-participating sub-catchments the data do not have the quality that will allow for more than a naïve qualitative comparison.
DATA REQUIREMENT ISSUES			
Type of data	Spatial data land use, topographic data, weather data and soil data	Data are available for the calculation of the indicator and creation of comparison groups.	The quality of the data for the modelling are suitable only for analysis at sub-catchment level, and not for within sub-catchment (field/farm level) analysis.
Primary monitoring data	No primary data used		
Sample size			
Spatial dimension	Indicator is calculated by sub-catchment		
Temporal dimension	Frequency of the data are dependent on the source of the data, Corine land cover data are not in sync with RDP programme cycle	Temporal land use data can be derived from IACS land use data which are annual data and these are summarised to broad land use classes.	IACS land use data are only available for agricultural land hence there are gaps in the data for a range of other land uses (forestry, semi-natural areas and urban development). CORINE data are used to fill these gaps.
Processing requirements	GIS software	Data processed in standard GIS using scripts	
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	Method used is based on robust and well documented theoretical model (USLE)	The quantitative indicator does not explain the effects	The IACS land use are not validated against Corine Land cover data or statistically tested for their accuracy

Assessment of net-impact	The estimation of direct and indirect effects needs the availability of control groups	Modelled changes for the indicator were used for DiD analysis.	The DiD is using the mean of the indicator value by sub-catchment and is unable to explain the participation and non-participation.
Micro-macro linkage	Micro level (25m raster cell) is embedded in the method, because the location of a farm within a sub-catchments determines the actual impact on loss of soil through erosion. At the method calculates the inflow, retention and outflow of sediment by 25m raster cells.	Sub-catchment level can be considered as a micro level in relation to the European soil data which are available at NUTS3 level, i.e. the case study area. An aggregation of the results from sub-catchment to NUTS3 level can be validated against the EU level values of the indicator.	
ADDITIONAL COMMENTS			

Public good: Water quality - diffuse pollution, Finland (WQ-FI)

Indicator and Method tested: Nitrogen reduction (calculated GNB nitrogen) and Biophysical/Structural modelling

Dimensions	**Pre-conditions**	**Solutions**	**Limitations**
Integration of counterfactuals	Structural models are flexible in determining one or multiple counterfactuals. However, the environmental indicator must be quantitatively and causally linked to the unit of analysis to enable evaluation. This requires large databases and previous studies on causal linkages.	Exploiting existing models with new data requires data in the same format as the model was designed for. Updating the model to keep up with changes in the agri-environmental programme specifics.	Adding new data to models requires time and expertise to check model consistency. The model is specific to the type of data, changes in data sources may require rebuilding the model. Changes in agri-environmental programme require model updates.
DATA REQUIREMENT ISSUES			
Type of data	FADN data and transfer functions describing fertilizer application to run-off amounts.	Use transfer functions that best describe environmental conditions at the study area.	The transfer functions describing run-off are not spatially explicit, especially when used to describe an average farm.
Primary monitoring data	The model relies on FADN data, crop and labour price information for the evaluation period. No primary environmental monitoring data is needed when evaluation uses pressure indicator (fertilizer use).		FADN data access may be limited.
Sample size	Statistical estimation requires large sample sizes, in this application the number of observations was over 1 500, from 343 farms over a period of 10 years.		
Spatial dimension	Farm-level data used to create a representative crop farm. Essentially an spatial model.	Evaluation results are understood at the aggregate spatial level, in this case Southern Finland.	Decreasing spatial resolution requires additional data on smaller spatial resolution, or assumptions on behavioural similarity or difference at specific locations.
Temporal dimension	The model employs data from multiple years. Modelling the average behavioural response over the evaluation period requires data covering the whole evaluation period.	FADN sample is annually updated	

Processing requirements	Running the model requires translating the data into a specific format. The model may need recalibration with new data.	Time for inputting new data should be reserved.	Model recalibration may be time intensive.
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	The behavioural model builds on causal relationships between the measure and profit maximizing farmer behaviour. Estimating environmental impacts in addition to pressure indicators requires transfer functions.	Models are built on economic theories. Transfer functions from relevant literature are employed.	The models are limited by their capability of taking external factors into account to a varying extent.
Assessment of net-impact	The model produces net-impacts on an aggregate level.		The models are limited by their capability of taking external factors into account to a varying extent. Out of sample prediction always carries a risk of biased evaluation results.
Micro-macro linkage	The model uses micro-level data to construct a representative farm model. The representative farm is essentially a macro-level agent.		Local impacts may be difficult to assess if farm type distribution differs much spatially.
ADDITIONAL COMMENTS			

Public good: Water quality- diffuse pollution, Germany (WQ-DE)

Indicator and Method tested: Mineral N content in the soil in autumn (Nmin) and Pairwise comparison and Regression analysis (plot and water protection area level)

Dimensions	**Pre-conditions**	**Solutions**	**Limitations**
Integration of counterfactuals	The evaluation approach can be used for both “Qualitative and Naïve quantitative” and “Elaborate statistics-based” analysis.	A “Naïve” quantitative analysis was applied in a previous project at the institute. Matching of similar farms has still improved the robustness of results of the pairwise comparison, compared to results without any matching attempt.	It was planned to update the previous analysis with micro level data for recent years. However, newer data was only provided as aggregated data set which limited sound statistical analysis.
DATA REQUIREMENT ISSUES			
Type of data	- Statistical sampling requires monitoring data - Secondary data (e.g. Agricultural Census; CLC; FADN; LPIS; FSS)	Monitoring data is used that includes a variety of variables on the farm structure and history and environmental conditions.	
Primary monitoring data	Monitoring data at field level	Annual monitoring data on autumn Nmin values have been used.	Newer data was only provided as aggregated data set due to data protection reasons which limited sound statistical analysis.
Sample size	- Farm sample (e.g., FADN's field of observation) - Large samples needed to robustly apply pairwise comparison and regression analysis.	Large sample of Nmin values was available at micro level which enabled to conduct the analysis at sub-measure level.	For some sub-measures the sample size was too small to detect the environmental effects.
Spatial dimension	Data applied at the farm/parcel level, strictly dependent by sample size (FADN, CLC)	Spatial data is not used for the analysis.	
Temporal dimension	The update for data is periodic, dependent by the type of datasets and data sources, often not in sink with RDP program cycle	Annual data were used.	In recent years micro level data is aggregated at the level of the drinking water protection area and cannot be used for this kind of analysis.
Processing requirements		Data can be processed with common statistical software.	
METHODOLOGICAL ISSUES			

Establishment of robust causal relationships	The robust causal relationship is based on theoretically sound models and methods to show the link between agricultural farming practices related to soil and nutrient management and its effect on water quality.	Nmin indicator is used in the analysis which is based on well-documented, theoretically sound models and methods. Causal relationships have been quantitatively assessed through matching approach.	Limited information on farm structure and management data did not require the application of advanced matching techniques to further improve the analysis of causal relationships.
Assessment of net-impact	The estimation of direct and indirect effects needs the availability of control groups.	The comparison groups for participants and non-participants are tested for direct effects.	Indirect effects were not considered as some effects such as substitution, multiplier, and spill-over seem not to be relevant for this assessment. Deadweight effects could not be considered as panel data for the reference group was not available (only with-without analysis).
Micro-macro linkage	The aggregation of micro level data has to be statistically verified in terms of representativeness.	The aggregation to the level of the water protection areas has been conducted to construct a data set similar to the macro level data set to improve consistency between results at micro and macro level.	Statistical representativeness has not been verified. Only estimations of environmental impacts at macro level are possible.
ADDITIONAL COMMENTS			

Public good: Water quality - Germany (WQ-DE)

Indicator and Method tested: Gross nutrient balance (GNB) and Propensity Score Matching (farm level)

Dimensions	**Pre-conditions**	**Solutions**	**Limitations**
Integration of counterfactuals	Propensity score matching (PSM) can be used for an “Elaborate statistics-based” analysis	The “Elaborate statistics-based” analysis requires well defined samples with a sufficient amount of observations and a variety of variable. Application of Propensity score matching has improved the robustness of results.	Results of the analysis are not explicit/ robust due to limitations of the data quality and quantity.
DATA REQUIREMENT ISSUES			
Type of data	- Biophysical model requires a series of input data (consumption of fertilisers, Gross Input of manure and other Inputs) - Potential surplus of nitrogen (GNS) on agricultural land and potential surplus of phosphorus on agricultural land (kg /ha/year) - Secondary data (e.g. Agricultural Census; CLC; FADN; LPIS; FSS)	Different data sources are used: - Monitoring data provided by the monitoring organization and managing authority - Control data of the fertilizer ordinance - Farm accountant data of the LAND-Data Ltd IACS data	Combination of data from different sources is challenging as structural differences do exists between the data sets.
Primary monitoring data	- Water use and fertilization input use - Monitoring data at farm level	Annual monitoring data on nutrient balances and farm structural data have been used.	Some data sources (e.g. control data of the fertilizer ordinance) do only include net nitrogen balances.
Sample size	- Farm sample (e.g., FADN's field of observation) - Large samples needed to robustly apply PSM.	Combinations of different data sources were explored to increase sample size.	Combination of data from different sources is challenging as structural differences do exists between the data sets.
Spatial dimension	Data applied at the farm/parcel level, strictly dependent by sample size (FADN, CLC)	Spatial data is not used for the analysis.	
Temporal dimension	- The update for data is periodic, dependent by the type of datasets and data sources, often not in sink with RDP program cycle - Gross Nutrient Balance (4 year average)	Annual data were used.	Data of controls of the fertilizer ordinance control different farms each year. Thus, panel data is not available.
Processing requirements	GNB is calculated as the balance between	Data can be processed with common statistical	

	inputs and outputs of nutrients to the agricultural soil	software.	
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	The robust causal relationship is based on theoretically sound models and methods to show the link between agricultural farming practices related to soil and nutrient management and its effect on water quality.	The CMEF impact indicator is used for the impact assessment of AEMs which is well-known and widely used for monitoring water quality. Causal relationships have been quantitatively assessed through propensity score analysis.	
Assessment of net-impact	The estimation of direct and indirect effects needs the availability of control groups.	The comparison groups for participants and non-participants are tested for direct effects.	Indirect effects were not considered as some effects such as substitution, multiplier, and spill-over seem not to be relevant for this assessment. Deadweight effects could not be considered as panel data for the reference group was not available (only with-without analysis).
Micro-macro linkage	Upscaling of micro level data has to be statistically verified in terms of representativeness.	External assumptions have been implemented to improve consistency between results at micro and macro level.	Estimations for the macro level could be done but a detailed location analysis is not possible due to limited data availability.
ADDITIONAL COMMENTS			

Public good: Water Quality- diffuse pollution, Greece (WQ-GR)

Indicator and Method tested: GNB+water use/ha and Biophysical model (land parcel and farm level)

Dimensions	**Pre-conditions**	**Solutions**	**Limitations**
Integration of counterfactuals	The biophysical model can be used for “Qualitative and Naive Quantitative” as well as "Elaborate statistics-based evaluation options" analysis.	A naive counterfactual approach was used comparing fields participating and non-participating in the AE action, using the available IACS geo-referenced data of 2011.	Given the lack of IACS data in different time points, no DiD approach could be applied.
DATA REQUIREMENT ISSUES			
Type of data	Biophysical model requires a series of input data (consumption of fertilisers, Gross Input of manure and other Inputs), potential surplus of nitrogen (GNS) on agricultural land	IACS georeferenced data, soil map, types of crop.	
Primary monitoring data	Water use and fertilization input use, monitoring data at farm level	Use of existing available data taking into account important crop types, soil conditions of the case study area in relation to the applied different farming practices of the AE action.	The Farm Accountancy Data Network data set lacks actual information on fertiliser application and/or water use
Sample size	Farm sample.	Biophysical model does not use farm samples but all land parcels that are supported by the examined measures (inventory).	
Spatial dimension	Data applied at the farm/parcel level, strictly dependent by sample size	The spatial dimension is based on the IACS field boundaries.	
Temporal dimension	The update for data is periodic, dependent by the type of datasets and data sources, often not in sink with RDP program cycle. Gross Nutrient Balance (4 year average)	Use of IACS geo-referenced data of 2011 available at land parcel level.	IACS georeferenced data are theoretically available every year. Monitoring water quality and quantity data are irregular.
Processing requirements	GNB is calculated as the balance between inputs and outputs of nutrients to the agricultural soil.	IACS georeferenced data were being processed using standard GIS software. The biophysical model calculated the amount of nitrogen applied, as well as GNB in the form of nitrogen losses per ha. Moreover, it provided information on the irrigation rates that should be applied in order to avoid groundwater overexploitation,	

		according to the national cross compliance rules, applicable in the area.	
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	Need to have well-documented, theoretically sound models and methods that show the links between farming practices and environmental outcomes	Data set and indicators were available at the land parcel level The biophysical model calculated the GNB in the form of nitrogen losses per ha and the water use/ha between participants and non-participants.	The obtained results were not verified with monitoring water quality and quantity data.
Assessment of net-impact	The estimation of direct and indirect effects needs the availability of control groups	The biophysical model provided quantifiable results.	IACS data constraints did not allow to apply a DiD counterfactual approach exploring changes between different comparison groups over time.
Micro-macro linkages	Macro level can build on micro level analysis	Micro and macro linkages considered only in an intuitive manner. Two macro level analysis 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. And the second on the actual distribution across soil classes.	Farm level which is the decision level for participation in the various schemes was missing. The functional unit was not linked to the any programmatic scale.
ADDITIONAL COMMENTS			

Public good: Animal Welfare, Germany (AW-DE)

Indicator and Method tested: Integration of animal-based indicators (result-based elements) in a multi-criteria framework for the evaluation of animal welfare impacts (micro level)

Dimensions	**Pre-conditions**	**Solutions**	**Limitations**
Integration of counterfactuals	The animal-based indicators can be used as part of an Naive quantitative" assessment using ad-hoc methods to consider sample selection issues as well as through statistics based analysis using explicit approaches for sample selection issues.	The statistics-based analysis needs well defined samples with sufficient number of observations either from secondary statistics or livestock monitoring data (farm visits) to perform robust matching methods such as propensity score matching at micro level.	Livestock monitoring data from farm visits are costly to gather. The available samples are thus only sufficient to apply naive quantitative analysis except in cases where sufficient monitoring data or secondary data sources exist and are accessible.
DATA REQUIREMENT ISSUES			
Type of data	The integration of animal-based indicators in a multi-criteria assessment of animal welfare requires data on livestock health issues (e.g. lameness, mortality and body condition), housing and management conditions. In addition the construction of comparison groups requires farm structural data.	The following data types were used: Livestock and herd data, livestock management data, and farm structural data.	The available data are not sufficiently exhaustive to build large samples covering different livestock and farm types. The availability of livestock data and data for non-participants restricts the use of robust counterfactuals.
Primary monitoring data	In addition to data available from existing livestock databases monitoring data from farm visits are required.	Livestock monitoring data were used.	See below under sample size
Sample size	Sample strategy of selected farms should cover a representative sample of different livestock and husbandry systems and include participating and non-participating to RDP. The sample needs to allow for the comparison of participant and non-participants with matched farm structural characteristics.	Livestock monitoring data from 150 dairy farms covering different farm characteristics were available for the case study.	Livestock monitoring through farm visits is very costly. In case of long term evaluation contracts different sampling strategies can be explored to collect primary data through farm visits.
Spatial dimension	Data applied at the farm level. Different farm types and livestock husbandry systems are key dimensions for animal welfare impacts.	Spatially-explicit factors were not considered in the design of the livestock monitoring (only farm structural aspects).	
Temporal dimension	Counterfactual based evaluation requires data	Secondary livestock and farm data are updated	Indicator can be influenced by seasonality,

	from at least two different years at the beginning and at the end of the RDP programming period.	annually. Empirical monitoring data from farm visits were only available for one point in time due to resource limitations.	which needs to be considered in the sampling strategy. The consideration of temporal dimensions largely relies on already existing livestock databases or livestock monitoring databases. In case of long term evaluation contracts different sampling strategies can be explored to collect primary data through farm visits at different points in time.
Processing requirements	Particular data processing tasks are the integration of primary and secondary data.	Empirical and secondary data were transformed into the unit of analysis of the indicators and plausibility and consistency checks were carried out with the data.	
METHODOLOGICAL ISSUES			
Establishment of robust causal relationships	Need to have well-documented, theoretically sound models and methods that show the links between measure prescriptions, changes in housing systems and livestock management and observed changes in different animal welfare criteria.	Conceptually and theoretically sounds models of the causal relationships could be developed for different relevant policy measures and animal welfare criteria and indicators.	The quantitative assessment of the causal relationship depends on the availability of bigger samples of livestock monitoring data as well as access to existing livestock data.
ADDITIONAL COMMENTS	The tested indicators add a direct (i.e. result-based) assessment of health criteria to the assessment of housing criteria through the use of resource or management based indicators. The tested indicators have a high acceptance by stakeholders and scientists and an application is recommended in combination with resource and management based indicators. The cost-effective application depends on available monitoring data. Few cases exist where livestock monitoring data are collected as part of animal welfare payments or some specific indicators are included in available livestock databases. High monitoring requirements and costs might prohibit the application if no data sources exist.		