ENVIEVAL

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

(Project Reference: 312071)

Area 2.1.4: Socioeconomic research and support to policies

KBBE.2012.1.4-08:

Report D8.1
Fact sheets of the tested indicators and methods

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

Approved by Work Package Manager of WP8: Gerald Schwarz (TI)

Date: November 2015

Approved by Project Coordinator: Gerald Schwarz (TI)

Date: November 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

Short introductory note to the fact sheets

The final outcome of the ENVIEVAL project is a handbook for the evaluation of environmental impacts of RDPs summarising the key characteristics, strengths and weaknesses of the different developed and tested indicators and evaluation methods in response to the main evaluation challenges. An important part of the handbook is the fact sheets for the different indicators and evaluation tools. The fact sheets are the final outcome of WP8 and provide a short summary of the main characteristics of the indicators and methods tested in the ENVIEVAL project. They provide information on why and for which policy aspects the indicators or methods can be used, and where the required data can be sourced and obtained. The fact sheets summarise the strengths and weaknesses of the indicators and methods, and highlight their contribution to addressing the main challenges. An adjusted 'SWOT' framework is used to synthesise the key advantages, disadvantages and contributions of the indicator / method.

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

Indicator fact sheets:

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

Method fact sheets:

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

The indicator fact sheets focus on additional non-CMES indicator tested in the ENVIEVAL project for their contributions to address indicator gaps in environmental evaluations of RDPs. The method fact sheets focus on advanced modelling approaches tested at micro and macro level for dealing with the complexity of public goods, considering other intervening factors and providing solutions for situations without (or very limited) non-participants. The fact sheets were reviewed by the stakeholder reference group and their comments and feedback integrated in the final version of the fact sheets. In addition, a selection of fact sheets was presented and discussed at the final project conference in Brussels on the 19th of November 2015.

Climate Stability

Method/Indicator: Carbon footprint

1. Definition / description of the method, including:
   - *Environmental public good*: Climate stability
   - *Micro or macro-level application*: Carbon Footprint (CF) is a well-established method to estimate carbon emission from functional units having different structural and management characteristics.
   - *Type of method/indicator*: Carbon Footprint (CF) can be considered as a method to quantify GHG net emissions as well as an indicator that measure these emissions. CF has been developed in the more general setting of 'ecological footprint' (EF) proposed by Rees (1992) for measuring the human 'load' considering the human carrying capacity as the maximum persistently-supportable load. EF could be considered as a composite indicator using either a common unit of measurement (e.g. the amount of productive land and sea area necessary to supply human population consumption) or an a-dimensional value system (irrespective of the measurement unit) such as the Agri-environmental Footprint Index proposed to evaluate agri-environment schemes (Purvis et al., 2009). The CF has been developed independently, in a modified hybrid form that derives only its name from EF, but conceptually is a global warming potential indicator developed through a specific method (Pandey and Agrawal, 2014). The carbon footprint approach allows us to measure the quantity of greenhouse gases (GHGs) expressed in terms of CO2 equivalent that is directly and indirectly caused by an activity or is accumulated over the life stages of a product and emitted into the atmosphere by an individual, organisation, process, product or event from within a specified boundary. The estimation of GHG footprint can be carried out by a process-based Life Cycle Assessment (LCA) where multiple environmental impact categories are assessed from cradle to grave. A very important element in the CF assessment is the functional unit considered and its system boundaries defined in temporal and physical terms that generally depend on the subject and the policy question (Minx et al., 2009). The functional unit could be the farm and/or the single productive process and its CF is the climate impact under a specified metric that considers all relevant emission sources, sinks, and storage in both consumption and production (Peters, 2010). The CF approach focuses on emission drivers, taking into account the indirect effects of farming practice changes on other sectors, e.g. on energy sector (changes on fuel consumption) or industry (changes on fertiliser and pesticide use).
   - *Reflected RDP priority and focus area*:
     - Priority 5 of the RD programmes: Promoting resource efficiency and supporting the shift toward a low-carbon and climate-resilient economy in agriculture, food and forestry sectors.

- Focus area 5D: Reducing greenhouse gas and ammonia emissions from agriculture; Focus area 5E Fostering carbon conservation and sequestration in agriculture and forestry.

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

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

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

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

5. Context of the case study testing
   - *Case study area*: Emilia Romagna Region of Italy

- *Policy context*: In the past RDP the main measures of reference were: Agri-environment sub-measures for climate change (214/A, 214/B and 214/E); Support for non-productive investments (216) and Increased renewable energy production (221).
- *Data used*: Ad-hoc survey for crops and livestock, where primary monitoring data related to land use and input use were collected by the evaluator. The IACS database will be used to distinguish participants to RDP programme from non-participants.
- *Evaluation approach tested*: The analysis of the GHG emission at process level has been carried out comparing different farming systems (organic and integrated vs. conventional ones), which have proven to be the best functional units to apply the CF methods. At farm level, the calculation of direct and indirect GHG emission can be done with the JRC Carbon Calculator (Bochu and Metayer, 2013), a user-friendly open-source tool designed to assess the life cycle of GHG emissions from different types of farming systems. The information contained in the FADN sample could be a good starting point, although additional information about farming practices is needed. CF allows for the creation of comparison groups of before-and-after as well as with-and-without participants. However, when the number of observations is insufficient for an elaborate statistics-based evaluation, a naïve group comparison counterfactual approach can be used, possibly assuring that sample selection has been minimised through expert knowledge to create similar comparable groups.

6. Strengths and weaknesses of the method (based on SWOT)

Evaluation challenges (relevant for methods)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	The use of site-specific data (FADN and direct surveys) allows taking account of specific local situations	Difficulty in considering matters that are strictly related to farm management (farming practices)	CF allows the measurement of RDP impacts at different levels
Timing of environmental impacts captured	Use of FADN data provides annual data by farms without the needs of specific environmental impact monitoring	Additional ad-hoc surveys are only done occasionally (e.g. RDP evaluation)	CF allows periodic assessment of the environmental performance of RDP measures
Establishment of robust causal relationships	CF allows a direct judgment of GHG performance of the single productive (functional) unit and is based on a theoretically sound basis	GHG emissions are calculated based on average CO_2 emission coefficients applied to individual farming systems, making them often unrepresentative in a local context. Data processing is quite time-intensive	An estimation of the overall effects of GHG emissions (including offsite emissions) is possible with the LCA approach
Assessment of net-impacts	The estimation of direct and indirect effects can be done for each of the control groups	FADN data allow only use of naïve methods for low number of observations for control groups	Process-level analysis could reduce the need for primary data monitoring
Establishment of consistent micro-	The method allows inference from the micro to the macro	Good upscaling depends on sample territorial	A statistically-sound representativeness of

Evaluation challenges (relevant for methods)	Strengths	Weaknesses	Key contribution to evaluation benefits
macro linkages	level (regional) by upscaling	representativeness and this is not always possible with process-based or FADN samples	ad-hoc surveys can ease the upscaling from farming system unit to regional level.
Appropriateness of indicator(s)	CF synthesises well the complexity of environmental relationships behind the GHG emissions		
Unambiguous and understandable results	The results can be easily communicated to target groups (managing authorities, policy makers and farmers)	The inner workings of calculating the CF are quite complex	

7. Recommended application
   - The estimation of GHG emissions and sinks is generally a time-intensive procedure which requires adequate expertise and could undermine its use due to very high costs for ad hoc monitoring surveys and data processing. Options to overcome this limitation can come from: a) the realisation of multi-purpose surveys on farming practices useful for more than one impact indicator; b) the design of software tools that could be a worthwhile investment beyond the monitoring and evaluation phase, increasing the use of these tools in the farmer decision-making process to improve performance efficiency.
   - Current existing databases such as FADN and IACS cannot provide all the types of required data for the carbon footprint calculation, lacking information on farming practices. Moreover, some problems may occur when the reference databases are not statistically representative about the quality and quantity of inputs purchased and used by farmers and the type of implementation of farming practices. Furthermore, the more are the variables to be considered in the production systems, the more CF analysis is complex (e.g., mixed farms compared to mono-cultural farming systems).
   - The lack of sector- and region - specific emission factors for important agricultural inputs add to the uncertainty. The standard method must address how to deal with alternative scenarios and land use changes.

References

Bochu J-L, Metayer N (2013) Development of Carbon Calculator to promote low carbon farming practices – User guidance manual for the Carbon Calculator, Deliverable to EC-JRC-IES by Solagro.

Minx JC, Wiedmann T, Wood R, Peters GP, Lenzen M, Owen A, Scott K, Barrett J, Hubacek K, Baiocchi G, Paul A, Dawkins E, Briggs J, Guan D, Suh S, Ackerman F (2009) Input–Output analysis and carbon footprinting: an overview of applications. *Economic Systems Research* 21(3): 187-216.

Pandey D, Agrawal M (2014) Carbon Footprint Estimation in the Agriculture Sector. In: Muthu, S.S. *Assessment of Carbon Footprint in Different Industrial Sectors*, Volume 1, EcoProduction.

Peters GP (2010) Carbon footprints and embodied carbon at multiple scales. *Environmental Sustainability* 2.

Purvis G, Louwagie G, Northey G, Mortimer S, Park J, Mauchline A, Finn J, Primdahl J, Vejre H, Vesterager JP, Knickel K, Kasperczyk N, Balazs K, Vlahos G, Christopoulos S, Peltola J (2009) Conceptual development of a harmonized method for tracking change and evaluating policy in the agri-environment: the Agri-Environmental Footprint Index. *Environmental Science & Policy* 12.

Rees, W. E. (1992) Ecological footprints and appropriated carrying capacity: what urban economics leaves out. Environment and Urbanisation 4 (2): 121–130.

Method: Sector models - DREMFIA

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

2. General requirements
   - *Data requirements*: As DREMFIA is an up-and-running sectoral model, all the data required for analysis are collected on a continual basis. Further, the complexity of the model requires multiple data sources, partly from official statistics and partly from other sources.
   - *Skill requirements*: Building up such a sector model as DREMFIA is time consuming, demanding and requires advantage skills. Also using the model needs trained personnel.

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

4. Context of the case study testing
   - DREMFIA results can be disaggregated from national to regional level. Regional effects are calculated for four main areas and 17 sub-areas. Farm-specific effects are not possible to assess with the model, but are also not as relevant due to the global nature of the assessed pressure indicator (i.e. CO_2 equivalent emissions).
   - As the majority of Finnish AEM sub-measures do not specifically target GHG emissions (except for the special measure for long-term grass/hay growing peat fields which has relatively few participants), we use the grouping of the AEM (214) and LFA (211, 212) for policy analysis (MMM, 2007). These two schemes affect the overall land use and production intensity – major contributors to agricultural CO_2 emissions – which the model essentially captures.
   - DREMFIA can construct counterfactuals without real-world comparison groups using a wide variety of data describing both domestic and international market conditions. Exogenously-determined EU prices influence domestic prices, but domestic prices may be different from EU prices. Four main areas are included in the model: Southern Finland, Central Finland, Ostrobothnia (the western part of Finland) and Northern Finland. Production in these areas is further divided into sub-regions on the basis of the support areas. In total, there are 17 different production regions. This allows a regionally disaggregated, exact description of policy measures and production technology.
   - DREMFIA uses multiple data sources to simulate the agricultural markets in Finland. The simulation model uses annual-level statistics collected between years 1995-2012 and also has its own collection of data. The used data sources are best represented as a list due to their number:
   - Data from official statistics used in simulation:
     - Prices of agricultural inputs, commodities and dairy products
     - Consumption of agricultural commodities and dairy products
     - Imports and exports of agricultural commodities and dairy products
     - Use of crops as fodder at farms and in fodder industry
     - Production yields per hectare and per animal
   - Data from official statistics used in model validation:
     - Agricultural total calculations on the value of different inputs in agriculture (similar to EEA)
     - Land use under different crops and number of animals at different regions and in the whole country
     - Farm structure statistics (FSS) – distribution of dairy cows in different farm size categories is endogenous in the model

- Data partially available in official statistics:
    - Agricultural payments (according to support regions and specific rules and definitions)
    - CAP pillar 1 payments
    - LFA-payments
    - Agri-environmental programme
    - National subsidies
    - Investment subsidies – specific to various kind of investments in livestock and crop production
- Other data:
    - Use of inputs in agricultural production - per ha, per head per year
    - Mainly from activity-based cost models maintained and published by national agricultural extension services (www.proagria.fi)
    - Partly from FADN – activity based unit cost calculations
    - Use of different feed stuffs per animal, from dairy farm recording system, and other livestock specific data systems of agricultural extension services (www.proagria.fi)
- Other knowledge used in simulation:
    - Specific needs of energy and protein content as well as roughage needs of different animals – Luke (Natural Research Insitute Finland) feeding norms
    - Nutrient contents of manure of different livestock
    - Luke internal calculations maintained in animal nutrition research and/or specific tables retrieved and summarised in different research projects
    - Nitrogen response function parameters
    - Milk yield response function parameters
    - Other technical parameters related to use of inputs per ha and head

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

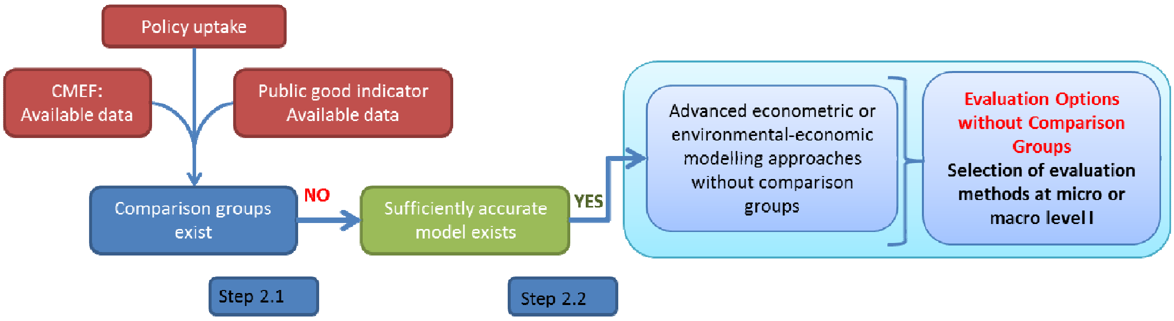

Figure 1 Evaluation steps of counterfactual logic model

5. Strengths and weaknesses of the method (based on SWOT)

Evaluation challenges	Strengths	Weaknesses	Key contribution to evaluation benefits
Compatibility with local environmental and farm structural characteristics		Regional level modelling does not consider local environmental characteristics.	
Timing of environmental impacts captured	The dynamic optimisation procedure has very high temporal resolution, and can show when effects are happening.		The modelling allows dynamic impact assessment also on the post-evaluation period.
Establishment of robust causal relationships	The model is well documented and transfer functions explaining environmental effects are based on relevant scientific literature. Policies are directly modelled with all their requirements, providing excellent grounds to understand how policy mixes work together (assuming profit maximisation).		
Establishment of consistent micro-macro linkages		The model does not incorporate farm level optimisation, rather a regionally representative farm, thus losing spatial resolution. However, increasing the spatial resolution would come at a great computational and design cost, and might not be feasible	
Assessment of net-impacts	The model can capture substitution effects.		The model allows for comparative repetition of the impact assessment with new data.
Appropriateness of method to capture complexity of	The complexity of the model allows for testing numerous counterfactuals and assumptions to		

Evaluation challenges	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
environmental relationships	see if the evaluated policy had an impact. For climate in particular, the model provides clear results on impacts. Other measurable environmental indicators include diffuse water pollution indicators, providing a chance to examine joint effects of policy.		
Unambiguous and understandable results	The results are quantitative and take into account the complex structure of the agricultural production and environmental effects in Finland. The results are highly useable in policy work, providing also a chance for ex-ante recommendations.		The model allows for a number of counterfactuals, thus allowing the policy makers to refine paths of development. The model also incorporates other public goods in the analysis, providing a chance for a more holistic impact assessment.

6. Recommended application

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

References

MMM (2007) The Rural Development Programme of continental Finland 2007-2013, approved 10.08.2007 (Manner-Suomen maaseudun kehittämisohjelma 2007 – 2013, Hyväksytty 10.8). In Finnish. http://www.maaseutu.fi/attachments/6BQPIuj8V/Manner-Suomen_maaseudun_kehittamisohjelma_2007-2013_090114.pdf. 375 p.

Soil Quality

Method: InVEST model

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

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

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

4. Context of the case study testing
   - *Case study area*: Aberdeenshire, Scotland
   - *Policy context*: Agri-environmental and forest-environment measures include objectives for maintaining soil carbon and avoiding soil loss through erosion
   - *Used data*: The LULC data are provided by IACS/LPIS geo-referenced summarised land-use data of 2008-2013, and national land-cover data (LCM2007) are used to fill any gaps in the IACS land-use data. Measure uptake data for soil-relevant measures (214, 223 and 225)

were used. Other data requirements are derived from additional local data sources (GB Ordnance Survey data, Scottish Soil data and expert knowledge).

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

5. Strengths and weaknesses of the method (based on SWOT)

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

Evaluation challenges	Strengths	Weaknesses	Key contribution to evaluation benefits
Appropriateness of indicator(s) to capture complexity of environmental relationships	The model does incorporate the complex environmental and spatial relationships in its calculation of the indicators.	The obtained results were not verified with monitoring soil quality data.	The biophysical model provides results based on site-specific environmental conditions enabling an impact assessment in the absence of suitable soil monitoring data.
Unambiguous and understandable results	Results are easy to communicate to laypersons.		Method provides user-friendly outcomes in the form of maps.

6. Recommended application

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

References

Sharp R, Tallis H, Ricketts T, Guerry A, Wood S, Chaplin-Kramer R, Nelson E, Ennaanay D, Wolny S, Olwero N, Vigerstol K, Pennington D, Mendoza G, Aukema J, Foster J, Forrest J, Cameron D, Arkema K, Lonsdorf E, Kennedy C, Verutes G, Kim CK, Guannel G, Papenfus M, Toft J, Marsik M, Bernhardt J, Griffin R, Glowinski K, Chaumont N, Perelman A, Lacayo M, Mandle L, Hamel P, Vogl A, Rogers L, Bierbower W (2015) InVEST +VERSION+ User's Guide. The Natural Capital Project, Stanford University, University of Minnesota, The Nature Conservancy, and World Wildlife Fund.

Water Quality

Method: Biophysical model

1. Definition / description of the method, including:
    - *Type of method:* Biophysical model
    - *Environmental public good:* Water quality
    - *Micro or macro level application:* Land parcel level. The availability of data at parcel level will allow the aggregation to the upper and measure level.

2. General requirements
    - *Data requirements*: Water use and fertilisation input use, monitoring data at farm level
    - *Skill requirements*: Spatial analytical /GIS skills

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

4. Context of the case study testing
    - *Case study area*: Karditsa regional department, Thessaly Plain, Greece
    - Policy context: AE action for the reduction of nitrate pollution caused by agriculture in NVZs
    - *Used data*: IACS geo-referenced data of 2011 for participants and non-participants including the number of hectares of supported area and type of crop, a soil map of the specific site of the plain area of Karditsa.
    - *Evaluation approach tested* (short explanation of the main logic model steps): Land parcel level and the specific site of the NVZ of Karditsa were the units of analysis for the micro and macro level, respectively. Given that IACS geo-referenced data were only available for the year 2011, two comparison groups were constructed. Thus, a naive counterfactual approach was used comparing only participants in the nitrate pollution reduction scheme and non-participants focusing on land potentially irrigated and cultivated under intensive crops.

5. Strengths and weaknesses of the method (based on SWOT)

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

6. Recommended application

   The applied biophysical method calculates the nitrogen fertiliser application, nitrogen loss and water use/ha for specific crops taking into account the soil texture, relief and the nitrogen balance equation. Therefore this model is recommended in cases where actual data on water quality and quantity, i.e. fertiliser and water use, are missing. The required data includes IACS/LPIS data between participants and non-participants at different points in time and a detailed soil map. Moreover, when data with geo-referenced information is accessible, this spatially explicit approach has a great potential improving the causality linkages.

Reference

Action Plan for Thessaly Plain (2001) Join Ministerial Decision No 25638/2905 (Official Journal Government of Greece, 1422B)

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

1. Definition / description of the indicator, including:
   - *Environmental public good*: Water quality
   - *Type of indicator*: Complementary result indicator. The autumn Nmin value provides information on the amount of nitrogen in the soil that is potentially polluting the groundwater due to leaching during winter.
   - *Reflected RDP priority and focus area*:
     - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
     - Focus area 4B: Improving water management, including fertiliser and pesticide management.
   - *Unit of measurement*: Amount of nitrogen per ha (kg N/ha) in the form of nitrate or ammonia, in soil depths between 0 and 90 cm.
   - *Type of data required*:
     - Soil samples of the mineral nitrogen content of soil in autumn (monitoring data) of participating and non-participating sites with a sufficient sample size (minimum 100 samples per (sub-) measure).
     - Farm structural data including information on site specific conditions (e.g. sink or source characteristics, soil type), farm management practices (type of crops, type of grassland-use, livestock density) and weather conditions.
   - *Scale and level of application*: parcel level (spatial)

2. Existing data sources
   - *EU-level*: Data is not available at EU level
   - *MS and regional level* (examples): In Lower Saxony, Germany, the data is collected for monitoring purposes by the managing authority. It is usually analysed at the level of the drinking water extraction areas.
   - *Fact sheets and information available from other sources*: A description of the indicators' characteristics and their application for impact assessment of AE measures was published by the monitoring organisation of Lower Saxony but is only available in German. (NLWKN, 2015 and NLWKN, 2010).

3. Context of the case study testing
   - *Case study area*: Lower Saxony, Germany
   - *Policy context:* Water protection measures to reduce diffuse pollution from agriculture are a key policy objective for agri-environmental policies.
   - *Used data:* Monitoring data of roughly 20,000 soil samples for the years 2000 to 2006 were used for micro-level analysis. For the years 2008-2012, only data aggregated at the level of drinking-water protection areas (i.e. important ground water areas) were available for the analysis.

- *Evaluation approach tested*: Nmin results for soil samples that were collected by the managing authority for monitoring purposes were used for the analysis. Two comparison groups were used to compare sites with AEM participation with non-participants. Samples of sites with similar environmental conditions were matched with each other. One site with AE measure was compared to three sites of non-participants. A pairwise comparison and regression analysis were conducted at micro level. The analysis at macro level used the aggregated data set that was provided by the managing authority for the recent years.

4. Strengths and weaknesses of the indicator (based on SWOT)

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

5. Recommended application
   - Autumn Nmin value can be used as a complementary result indicator for the evaluation of measures and sub-measures at parcel level. The timing of the measurement of the Nmin indicator is very close to the implementation of the AE measures. Thus, it is well suited for the annual impact assessment of water protection measures on agricultural land.
   - Basic sampling requirements for a robust impact assessment:
     1. Suitable site-specific conditions
     2. A minimum of 16 punctures per area
     3. Sample taking from October to mid-November (before the leakage water formation)
     4. If precipitation in autumn is high, sampling depth has to be adapted (deeper than 90 cm) (NLWKN, 2015)
   - Timing of sampling can reduce risk of bias due to climatic conditions.
   - The suitability of the indicator for statistics-based approaches (e.g. such as propensity score matching) to consider sample selection issues depends on the availability of, and

access to, sufficient annual monitoring and farm structural data. It is recommended to use the indicator in combination with the CMES impact indicator GNB which is well-known and widely used for monitoring water quality.

References

NLWKN (2015) Anwenderhandbuch für die Zusatzberatung Wasserschutz Grundwasserschutzorientierte Bewirtschaftungsmaßnahmen in der Landwirtschaft und Methoden zu ihrer Erfolgskontrolle. Grundwasser, Band 21.

NLWKN (2010) Untersuchung des mineralischen Stickstoffs im Boden. Empfehlungen zur Nutzung der Herbst-Nmin-Methode für die Erfolgskontrolle und zur Prognose der Sickerwassergüte. Grundwasser, Band 8.

Indicator: Water use/ha

1. Definition / description of the indicator, including:
   - *Environmental public good*: Water quality
   - *Type of indicator*: The new CMES impact indicator is water abstraction in agriculture (European Commission, 2013). This indicator refers to the volume of water which is applied to soils for irrigation purposes. Data concern water abstraction from total surface and ground water.
   - *Reflected RDP priority and focus area*:
     - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
     - Focus area 4B: Improving water management, including fertiliser and pesticide management.
   - *Unit of measurement*: m3 of water abstracted for irrigation
   - *Type of data required*: The indicator of water use/ha is estimated by a biophysical model using the applied water per dose, a detailed soil map and type of crop. LPIS-GIS data for participants and non-participants at different points in time are needed. Since IACS-LPIS data are available at different points in time, the deadweight effects could also be estimated.
   - *Scale and level of application*: The indicator on water abstraction could be calculated at NUTS 2 level ideally (and River Basin level); an analysis at regional level is more appropriate to capture the effects and impacts of the CAP on the environment.

2. Existing data sources
   - EU-level: The Survey on Agricultural Production Methods (SAPM) provides estimates of water use for irrigation on farm level. SAPM is a unique survey carried out by Eurostat in 2010 to collect data at farm level on agri-environmental measures. Data on water abstraction for irrigation cannot yet provide a pan-EU coverage (Eurostat, Water abstraction).
   - MS and regional level (examples): Annual data available for the period 1970-2009 depending on availability for each MSs (In 2007, 2008, 2009 data are available for 19, 11, 10 MSs respectively, Eurostat/OECD Joint Questionnaire).
   - Fact sheets and information available from other sources:
     1. IRENA, Indicator Fact Sheet 34.3-Share of agriculture in water use
     2. IRENA Indicator Fact Sheet 22-Water abstraction

3. Context of the case study testing
   - *Case study area*: Karditsa regional department, Thessaly Plain, Greece
   - *Policy context*: AE action for the reduction of nitrate pollution caused by agriculture in Nitrate Vulnerable Zones (NVZ)

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

4. Strengths and weaknesses of the indicator (based on SWOT)

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

5. Recommended application

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

References

European Commission (2013) Impact Indicators, Draft – Work in Progress for Discussion in the expert group on monitoring and evaluating the CAP.
http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetailDoc&id=6707

Eurostat, Water abstraction. http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_water_abstraction

IRENA, Indicator Fact Sheet 34.3-Share of agriculture in water use. http://ec.europa.eu/eurostat/documents/2393397/2518916/IRENA+IFS+34.3+-+Share+of+agriculture+in+water+use_FINAL.pdf/78932951-7252-4de2-8998-caab964c4246

IRENA Indicator Fact Sheet 22-Water abstraction. http://ec.europa.eu/eurostat/documents/2393397/2518916/IRENA+IFS+22.pdf/c600296a-d688-446a-8d71-7f5fa05483d3

Indicator: Gross Nitrogen Balance

1. Definition / description of the indicator, including:
   - *Environmental public good*: Water quality
   - *Type of indicator*: Gross Nitrogen Balance is an impact indicator proposed by the CMES that indicates potential surplus of nitrogen on agricultural land.
   - *Reflected RDP priority and focus area*:
     - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
     - Focus area 4B: Improving water management, including fertiliser and pesticide management.
   - *Unit of measurement*: kg of nitrogen per ha per year
   - *Type of data required*: Gross Nitrogen Balance is estimated by a biophysical model using the nitrate fertiliser usage per land parcel, a detailed soil map and type of crop. LPIS-GIS data for participants and non-participants at different points in time are needed. Since IACS/LPIS data at different points in time are available, the deadweight effects could also be estimated.
   - *Scale and level of application*: from land parcel to regional and national level.

2. Existing data sources
   - *EU-level*: Data for EU-27 could only be compiled for 2005-2008 (Eurostat, GNB)
   - *MS and regional level (examples)*: Gross nitrogen balances are not comparable between countries due to differences in definitions, methodologies and data sources used by countries. Nitrogen surplus (kg N/ha) is available for Norway and Switzerland between 1990 and 2008 (Nitrogen outputs, (kg N per ha), 1990-2008,EU-27, CH and NO)
   - *Fact sheets and information available from other sources*:
     - RDP 2007-2013, CMEF, Guidance note J – Impact Indicator Fiches, Improvement in water quality
     - IRENA indicator fact sheet 18.1 Gross nitrogen balance

3. Context of the case study testing
   - *Case study area*: Karditsa regional department, Thessaly Plain, Greece
   - *Policy context*: AE action for the reduction of nitrate pollution caused by agriculture in Nitrate Vulnerable Zones (NVZ)
   - *Used data*: IACS geo-referenced data of 2011 for participants and non-participants including the ha of supported area and type of crop, a soil map of a specific site of the plain area of Karditsa
   - *Evaluation approach tested*: Land parcel level and the specific site of the NVZ of Karditsa were the units of analysis for the micro and macro level, respectively. Given that IACS georeferenced data for participants and non-participants were only available for 2011, 2 comparison groups were constructed, using a naive counterfactual approach comparing only participants in the nitrate pollution reduction scheme and non-participants.

4. Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Using the GNB in combination with a detailed soil map enables the consideration of specific environmental characteristics, such as soil texture and relief, as well as significant crops.	Missing information on farm characteristics (e.g. use of agricultural area, type of farming system, participation in RDP or other policy measures) can restrict the application of the indicator in elaborate statistics-based evaluations.	The impact of the AE action is estimated within each soil class.
Appropriateness of indicator(s) to capture complexity of environmental relationships	The GNB is strongly related to the nitrogen pollution from agricultural sources.	Actual data on fertiliser inputs are missing.	The impact of the AE action takes into account the different farming practices applied.

5. Recommended application

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

References

Eurostat, Gross Nitrogen Balance. http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_gross_nitrogen_balance

Nitrogen outputs (kg N per ha), 1990-2008,EU-27, CH and NO. http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Nitrogen_outputs_%28kg_N_per_ha%29,_1990-2008,_EU-27,_CH_and_NO.png

CMEF, Guidance note J – Impact Indicator Fiches, Improvement in water quality. http://ec.europa.eu/agriculture/rurdev/eval/guidance/note_j_en.pdf

IRENA Indicator fact sheet 18.1 Gross Nitrogen Balance. https://circabc.europa.eu/sd/a/09cc1348-8232-447e-ae03-87e71d680b08/IRENA%2018.1%20-%20Gross%20nitrogen%20balance.pdf

Indicator: GNB for the assessment of effects of advisory services

1. Definition / description of the indicator, including:
   - *Environmental public good*: Water quality
   - *Type of indicator*: CMES impact indicator
   - *Reflected RDP priority and focus area*:
     - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
     - Focus area 4B: Improving water management, including fertiliser and pesticide management.
   - *Unit of measurement*: amount of nitrogen per ha utilized agricultural area (kg N/ha)
   - *Type of data required*:
     - Gross nitrogen balances for participating and non-participating farms that are calculated equally and reliably (to ensure comparability)
     - Farm structural data, particularly on land-use such as share of grassland and main crops, and livestock density
     - Reliable information on nitrogen cycles of farms to calculate robust gross nutrient balances (including import and export of organic N and purchases of feed and seed)
   - *Scale and level of application*: nitrogen balances are calculated at farm level

2. Existing data sources
   - *EU-level*: Data for EU-27 could only be compiled for 2005-2008 (update is planned for July 2016). As methodologies (especially with regards to the coefficients) and data sources used in different countries vary substantially, the balances are not consistent across countries, which means that data cannot be compared between countries (http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_gross_nitrogen_balance). To improve the comparability, harmonisation of methods and data sources in the member states is essential. The current situation also results in different qualities of evaluations due to different methods of calculations. Minimum requirements for data quality and the analysis should be defined to achieve comparability of good quality nitrogen balances.
   - *MS and regional level* (examples): Lower Saxony, Germany: Nitrogen balances are used to analyse the effect of advisory services (in combination with AE measures) on water quality. Database includes nitrogen balances of participating farms and a reference group generated from the controls of the fertiliser ordinance. An additional data set of 160 model farms was established in target areas of the Water Framework Directive (WFD). Recently, a reference group with farms outside of the WFD target areas was created. A description of the indicator (in German) is included in a handbook of the managing authority (NLWKN, 2015a)
   - *Fact sheets and information available from other sources*: Water quality impact indicator fiche by EU-Commission:

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

3. Context of the case study testing
   - *Case study area*: Lower Saxony, Germany
   - *Policy context*: extension services addressing the protection of water resources are very important. Advisory services of farmers are supported and an own measure in drinking water protection areas was established. The main measures of RD programmes are: 1) 323 Rural heritage (support of technical advice in drinking water protection areas) and 2) 114 Use by farmers and forest holders of advisory services (farm management: focus on nutrient management).
   - *Used data*:
     - Data from 160 model farms that receive intensive advisory service due to their location in target areas of Water Framework Directive are available for the years 2006 to 2012. As a reference group, farms of the fertiliser ordinance controls without any AE measure or advisory service are used.
     - The data set of the controls of the fertiliser ordinance contains farms participating in advisory services related to drinking water protection. Farms that are not participating in this measure or in AE measures are used as the reference group.
   - *Evaluation approach tested*: Nitrogen balances are compared between participants and non-participants of advisory services related to the improvement of water quality. The classic approach is used with two comparison groups for each analysis. Non-participants are compared with farms that receive advisory services. For the first test (combination of model farms with data of the controls of the fertiliser ordinance), two data sets that stem from different sources are combined. Thus, before conducting comparative statistical analysis, the two data sets have to be tested for structural differences. It turned out that the comparability of the two different data sets is limited. Recently, a reference group for the model farms was constructed and a comparative analysis was conducted by the monitoring organisation (NLWKN). Results show reduction effects of advisory services on the nutrient balances of farms with advisory service over the years while the nutrient balances of non-beneficiaries remained static (NLWKN, 2015b). However, this data was not available at the time of the analysis. Thus, a comparison of participants with non-participants using the control data set of the fertiliser ordinance was conducted. Propensity Score Matching is used to match similar farms. As gross nutrient balances are not included in the data set of the fertiliser ordinance controls, net nutrient balances are used for the comparative analysis.

4. Strengths and weaknesses of the indicator (based on SWOT)

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

5. Recommended application
   - GNB indicator provides reliable information on nitrogen management of farms and is widely used for the analysis of nutrient surpluses and impact assessment of AE measures due to its explanatory power (NLWKN, 2015a)
   - GNB indicator can also be used to assess the effects of advisory services related to water quality using a similar approach as for the analysis of AE measures
   - Net-nutrient balances are based on rough estimations on the amount of organic N as well as estimations of forage and grassland yields. Therefore, gross nutrient balances are more reliable and should be favoured over net balances.
   - To enable the application of advanced statistics based assessments of net-effects data should include sufficient information on gross nutrient balances and farm structure and management practices for participating and non-participating farms. Minimum requirements are:
     - Gross nutrient balances should be available for both groups
     - Reliable information on the components of the Gross Nitrogen Balance which cover the whole nitrogen cycle of the farm
     - Information on main crops and grassland share as well as livestock density
     - Information on type of AE and other RD measure

- Panel data for participants and non-participants should be available
- In addition to the nitrogen balances, single components of the balance, e.g. the amount of mineral fertiliser purchased, can also be used for impact assessment (NLWKN, 2015a)

References

NLWKN (2015a) Anwenderhandbuch für die Zusatzberatung Wasserschutz Grundwasserschutzorientierte Bewirtschaftungsmaßnahmen in der Landwirtschaft und Methoden zu ihrer Erfolgskontrolle. Grundwasser, Band 21.

NLWKN (2015b) Erfolgskontrolle von Grundwasserschutzmaßnahmen mit Hoftorbilanzen eines Referenzbetriebsnetzes außerhalb der Trinkwassergewinnungsgebiete und der WRRL-Beratungskulisse. Grundwasser, Band 25.

Landscape

Method: Spatial analysis

1. Definition / description of the method, including:
   - *Type of method*: Spatial analysis
   - *Environmental public good*: Landscape
   - *Micro or macro level application*: The availability of representative data at the farm level would allow for aggregation at the macro level.

2. General requirements
   - *Data requirements*: UAA (IACS, FADN, etc.), land cover/land use data, remote sensing and aerial photography data, landscape and vegetation maps
   - *Skill requirements*: Spatial analytical /GIS skills.

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

4. Context of the case study testing
   - *Case study area*: Northern Santorini, Greece
   - *Policy context*: 1.) Agri-environmental measure for the landscape protection: action for the maintenance of pruning and propagation practices in vineyards on the island of Santorini, 2.) Special measures in favour of the Small Aegean Islands concerning certain agricultural products, such as the continued cultivation of traditional vines.
   - *Used data*: Land-cover data of Northern Santorini drawn from GE images (2003 and 2012), IACS geo-referenced data of 2011 for participants.
   - *Evaluation approach tested*: Given that land-cover data were available between two periods in time (2003 and 2012), while IACS data provided only data for participants in 2011, three comparison groups were constructed, participants in the AE action, participants in the AE action plus the special aid to the Aegean islands, and participants in neither scheme.. A conceptual DiD approach was selected comparing the changes in traditional vineyards among the three comparison groups from 2003 to 2012.

5. Strengths and weaknesses of the method (based on SWOT)

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

6. Recommended application

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

References

Fonji SF, Taff GN (2014) Using satellite data to monitor land-use land-cover change in North-eastern Latvia. *SpringerPlus* 3: 61.

Hu Q, Wu W, Xia T, Yu Q, Yang P, Li Z, Song Q (2013) Exploring the use of Google Earth imagery and object-based methods in land use/cover mapping. *Remote Sensing* 5(11): 6026-6042.

Method: Landscape metrics

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

2. General requirements
   - *Data requirements*: agricultural land use data (IACS, LPIS), land cover data (CORINE land cover or national equivalents). The method is sensitive to data scale; hence data in an assessment need to be of compatible resolutions for a comparison of results.
   - *Skill requirements*: spatial and statistical analytical skills, sound GIS skills.

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

4. Context of the case study testing
   - *Case study area*: Grampian region, Scotland
   - *Policy context*: During the RDP 2007 – 2013 AE actions aimed to safeguard and enhance landscape; native woodland; non-native woodland; and geo-diversity. The main measures were: agri-environment (214), woodland creation (223) and woodland management (225).
   - *Used data*: IACS geo-referenced land-use data for 2008-2014 for participants in measures 214, 223 and 225 and non-participants, Land Cover Map 2007 (CEH, 2011) to fill data gaps, and the National Forest Inventory (Forestry Commission).
   - *Evaluation approach tested*: Land-use patches and the case study area as a whole were the unit of analysis for the micro and macro level, respectively. The data allowed for the creation of comparison groups of before and after as well as with and without participation. However the data were insufficient to explain the participation for an elaborate statistics-based evaluation; instead a naïve Difference-in-Difference counterfactual approach was used.

5. Strengths and weaknesses of the method (based on SWOT)

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

6. Recommended application

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

References

Botequilha Leitao A, Miller J. Ahern J, McGarigal K (2006) *Measuring Landscapes – a planner's handbook*. Island Press, London.

Farina A (2007) *Principles and Methods in Landscape Ecology – towards a science of landscape*. Springer Landscape Series 3.

Indicator: Shannon Diversity Index

1. Definition / description of the indicator, including:
   - *Environmental public good*: Landscape
   - *Type of indicator*: Shannon Diversity Index (SDI) is a proposed additional impact indicator. The SDI is most commonly used for the assessment of ecological diversity; however it is also applied for the assessment of landscape diversity.
   - *Reflected RDP priority and focus area*:
     - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
     - Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.
   - *Unit of measurement*: proportion of landscape occupied by patch type
   - *Type of data required*: land use/land cover data, IACS/LPIS
   - *Scale and level of application*: local, regional and national areas.

2. Existing data sources
   - *EU-level*: Data for EU-27 is available through the CORINE land cover data for 2006 only (http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3).
   - *MS and regional level* (examples): at national level land cover data may have local modifications but generally these can be reclassified into CORINE classes to facilitate comparison between MS. For the UK this is Land Cover Map 2007 (LCM2007).
   - *Fact sheets and information available from other sources*: none, but linked to the EU landscape convention.

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

4. Strengths and weaknesses of the indicator (based on SWOT)

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

5. Recommended application

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

References

Botequilha Leitao A, Miller J. Ahern J, McGarigal K (2006) *Measuring Landscapes – a planner's handbook*. Island Press, London.

McGarigal K (2015) Fragstats Help (http://www.umass.edu/landeco/research/fragstats/documents/fragstats.help.4.2.pdf)

Indicator: Patch shape index

1. Definition / description of the indicator, including:
   - *Environmental public good*: Landscape
   - *Type of indicator*: patch shape index is a proposed additional programme specific result indicator, which measures the geometric complexity of a patch (i.e. an area of the same land use/cover) and the impact of RDP on the landscape.
   - *Reflected RDP priority and focus area*:
     - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
     - Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.
   - *Unit of measurement*: ratio of patch perimeter (m) and the square root of patch area (m^2), adjusted by a constant to adjust for a square standard.
   - *Type of data required*: land use/land cover data, IACS/LPIS
   - *Scale and level of application*: patch and landscape level

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

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

4. Strengths and weaknesses of the indicator (based on SWOT)

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

5. Recommended application

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

References

Botequilha Leitao A, Miller J. Ahern J, McGarigal K (2006) *Measuring Landscapes – a planner's handbook*. Island Press, London.

McGarigal K (2015) Fragstats Help (http://www.umass.edu/landeco/research/fragstats/documents/fragstats.help.4.2.pdf)

Indicator: Land-cover change

1. Definition / description of the indicator, including:
   - *Environmental public good*: Landscape
   - Type of indicator: Additional programme specific indicator. Moreover, land-cover change is a pressure indicator based on IRENA operation (No24) that identifies land-cover changes to and from forest/semi-natural and agricultural land (EEA Report, 2006).
   - *Reflected RDP priority and focus area*:
     - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
     - Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.
   - *Unit of measurement*: Changes in land cover classified by type and size (%)
   - *Type of data required*:
     - LPIS-GIS and land-cover data between participants and non-participants before and after measure implementation at land parcel level.
     - Also, including information on farm structural variables (e.g. use of agricultural area, yields, type of farming system, input cost, participation in RDP or other policy measures) will enable you to assess the indirect effects, such as the deadweight effects, by checking change/maintenance observed in non-participants that would have occurred even in the absence of the applied measures.
   - *Scale and level of application*: spatial and temporal scale, NUTS 2/3 level (where data is available).

2. Existing data sources
   - *EU-level*: Since 2006, EUROSTAT has carried out a survey on the state and the dynamics of changes in land use and land cover in the European Union; this is called the LUCAS survey (Land Use/Cover Area Frame Statistical Survey). LUCAS is a field survey based on an area-frame sampling scheme. Data on land cover and land use are collected, and landscape photographs are taken to detect any changes to land cover/use to European landscapes. These surveys are done every three years.
   - *MS and regional level* (examples): There are two main types of information derived from LUCAS: aggregated statistical data and elementary data (for individual survey points). The aggregated results show land cover and land use for the EU-27 and national averages for the EU Member States, and can also be shown at a more detailed level, for example, for more than 250 NUTS 2 regions. Moreover, relevant work published by WWF-Greece presents the spatial data on land cover and its change tendencies in Greece from 1987 to 2007. (http://www.wwf.gr/en/areas/forests/land-uses).
   - *Fact sheets and information available from other sources*: None

3. Context of the case study testing
   - *Case study area*: Northern Santorini, Greece
   - Policy context: 1. AE action for the maintenance of pruning and propagation practices in vineyards on the island of Santorini, 2. Special measures in favour of the Small Aegean Islands.
   - *Used data*: Land-cover data of Northern Santorini drawn from GE images (2003 and 2012), IACS geo-referenced data of 2011 for participants in the AE action and the special measure.
   - *Evaluation approach tested*: Land parcel level and landscape of Northern Santorini were the units of analysis for the micro and macro level respectively. Land-cover data were available between two periods in time (2003 and 2012), while IACS data provided only for participants of 2011. Due to data availability, three comparison groups were constructed, participants in the AE action, participants in the AE action and the special aid to the Aegean islands and participants in neither scheme. A conceptual DiD approach is selected comparing the changes in traditional vineyards among the three comparison groups from 2003 to 2012.

4. Strengths and weaknesses of the indicator (based on SWOT)

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

5. Recommended application

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

References

European Environment Agency (2006) Report on Integration of environment into EU agriculture policy - the IRENA indicator-based assessment report. (http://www.eea.europa.eu/publications/eea_report_2006_2/downloadEEA report, 2006)

Land cover and land use (LUCAS) statistics. http://ec.europa.eu/eurostat/statistics-explained/index.php/Land_cover_and_land_use_%28LUCAS%29_statistics

Greece Then and Now: Longitudinal Mapping of Land Cover, 1987–2007, WWF Greece. http://www.wwf.gr/en/areas/forests/land-uses

Indicator: Visibility of change

1. Definition / description of the indicator, including:
   - *Environmental public good*: Landscape
   - *Type of indicator*: Visibility of change is a proposed alternative impact indicator which captures the changes in visibility of individual patches due to uptake of RDP measures.
   - *Reflected RDP priority and focus area*:
     - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
     - Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.
   - *Unit of measurement*: Area of visible land-use change; proportion of visible land-use change within landscape character area
   - *Type of data required*: digital terrain model, IACS/LPIS data, and topographic data,
   - *Scale and level of application*: From feature to regional, landscape character and national level

2. Existing data sources
   - *EU-level*: Data of relevant scale and spatial resolutions are not held at an EU level
   - *Member State and regional level* (examples): Relevant Member State data: IACS RDP uptake, spatial units (i.e. generally field boundaries), national topographic mapping (Ordnance Survey MastermapTM), Digital Terrain Model (1:10,000) and Landscape Character Assessment mapping.
   - *Fact sheets and information available from other sources*: No fact sheets are available, but further information on the indicator can be obtained from other scientific publications listed under references below.

3. Context of the case study testing
   - *Case study area*: Grampian region, Scotland.
   - *Policy context*: Agri-environmental and forest-environment measures aimed to safeguard and enhance the landscape and its character; native woodland and associated habitats and species; non-native woodland and associated habitats and species; and geo-diversity.
   - *Used data*: IACS geo-referenced data of 2009 and 2014 for participants and non-participants including the type of crop, National Forest Inventory, topographic data (Ordnance Survey MastermapTM) and Digital Elevation Model (DEM) data, Landscape Character Assessment.
   - *Evaluation approach tested*: The land parcel is the unit of analysis for the macro-level analysis. At Step 1, for the case study area, a baseline of the visibility of land-cover types is calculated to enable comparisons of before and after. Information on the types of features associated with the uptake of RDP measures is used as input for the land parcel. At Step 2, the macro level analysis is carried out with respect to the visibility of individual units, which are cumulated to landscape character areas. At Step 4, the outputs can be

presented as time series, and different groupings of land parcels selected to show impacts with and without uptake, stratified according to size, distribution (e.g. clustered or distributed according to a specified pattern geographically, or temporally), and interpreted with respect to landscape character map units.

4. Strengths and weaknesses of the indicator (based on SWOT)

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

5. Recommended application

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

References

Miller DR (2001) A method for estimating changes in the visibility of land cover. Landscape and Urban Planning 54: 93-106.

Sang N, Ode A, Miller DR (2008) Landscape metrics and visual topology in the analysis of landscape preference. Environment and Planning B 35: 504-520.

Indicator: Visual amenity

1. Definition / description of the indicator, including:
   - *Environmental public good*: Landscape
   - *Type of indicator*: Additional programme-specific indicator adapted by the research team for measuring the amenity values offered by the traditional vineyards. Given that there is a causal link established between Santorini's traditional vineyards and its contribution to making this landscape attractive for people, we assume that the alteration of iconic vineyards resulting from the characteristic pruning systems will deteriorate its distinctive landscape and consequently people's amenity will be affected negatively. The estimation of visual amenity indicator is categorised into three levels of an ordinal scale, high-medium-low. High level is considered when the pruning system is maintained, medium when land cover (vineyard) is maintained, low when the land cover is changed. The allocation of the visual amenity level to each comparison group is based on the observed changes in area of vineyard in the timeframe 2003-2012.
   - *Reflected RDP priority and focus area*:
     - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
     - Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.
   - *Unit of measurement*: loss of amenity values based on the observed land cover changes from 2003 to 2012 (%)
   - *Data requirements*: The specific indicator is based on land cover changes observed in area of traditional vineyards in the timeframe 2003-2012. Thus LPIS-GIS and land-cover data between participants and non-participants at different points in time are required.
   - *Scale and level of application:* spatial & temporal scale, from land parcel to regional level

2. Existing data sources
   - *EU-level*: -
   - *MS and regional level* (examples): -
   - *Fact sheets and information available from other sources*: -

3. Context of the case study testing
   - *Case study area*: Northern Santorini, Greece
   - *Policy context*: 1. AE action for the maintenance of pruning and propagation practices in vineyards on Santorini, 2. Special measures in favour of the Small Aegean Islands.
   - *Used data*: Land cover data of Northern Santorini drawn from GE images (2003 and 2012), IACS geo-referenced data of 2011 for participants.
   - *Evaluation approach tested*: Land parcel level and landscape of Northern Santorini were the units of analysis for the micro and macro level respectively. Given that land-cover data were available between two periods in time (2003 and 2012) and IACS data provided only for participants of 2011, three comparison groups were constructed, participants in the

AE action, participants in the AE action and the special aid to the Aegean islands and participants in neither scheme. A conceptual DiD approach was selected comparing the amenity values based on their observed land cover changes from 2003 to 2012 among the three comparison groups.

4. Strengths and weaknesses of the indicator (based on SWOT)

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

5. Recommended application

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

References

Arriaza M, Cañas-Ortega J, Cañas-Madueño J, Ruiz-Aviles P (2004) Assessing the visual quality of rural landscapes. *Landscape and Urban Planning* 69(1): 115-125.

Howley P, Donoghue C, Hynes S (2012) Exploring public preferences for traditional farming landscapes. *Landscape and Urban Planning* 104(1): 66-74.

Tahvanainen L, Ihalainen M, Hietala-Koivu R, Kolehmainen O, Tyrväinen L, Nousiainen I, Helenius J (2002) Measures of the EU Agri-Environmental Protection Scheme (GAEPS) and their impacts on the visual acceptability of Finnish agricultural landscapes. *Journal of Environmental Management* 66(3): 213-227

Biodiversity HNV

Indicator: High Nature Value forestry

1. Definition / description of the indicator, including:
   - *Environmental public good*: Biodiversity High Nature Value (HNV) forestry
   - *Type of indicator: The indicator shows the increase or decrease of ecotone length between afforested and adjacent land.*
   - *Reflected RDP priority and focus area:*
     - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
     - Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.
   - *Type of data required*: IACS georeferenced data, Forest cadastre data, Orthophoto image, land-use data.
   - *Scale and level of application*: Each land parcel of each relevant RDP measure at micro level and the group of land parcels of each relevant RDP measure in the geographical region at macro level.

2. Existing data sources:
   - *EU-level*: IACS data, orthophoto image, land-use data
   - *MS and regional level* (examples): IACS data is available through national rural agencies.
   - *Fact sheets and information available from other sources*: -

3. Context of the case study testing
   - *Case study area*: Nevezis wooded agrarian and urbanised plain area, Lithuania
   - *Policy context*: RDP measures: 214 Agri-environment payments, 221 First afforestation of agricultural lands, 223 First afforestation of non-agricultural lands. Since forest land is rapidly increasing on abandoned land, the ecotone[1] would be a good indicator to measure the afforestation measure implementation in the area.
   - *Used data*: IACS-LPIS georeferenced data, Forest cadastre database, Georeferenced spatial data set at 1:10,000 scale of the Republic of Lithuania (GDR10LT), Orthophoto image 2010-2014.
   - *Evaluation approach tested*: Using the before-after approach, the changes in the length of the ecotone was calculated in the selected area, to see what impact the application of RDP measures had on the heterogeneity of landscapes. At micro level, randomly-selected

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

land parcels (as accounted in LPIS databases) were chosen for the calculations to see if the applied measure has a positive or negative effect on the heterogeneity of landscapes. The problem with the application of the method at micro level was that it is not an automated process and it is time intensive. Every parcel had to be calculated and reviewed manually. At macro level, the method does not calculate the effect of each land parcel but the effectiveness of the measure as it showed the consolidated results from all the parcels in the region. The effectiveness was measured by extension or decrease of the ecotone length. This process was an automated one. The main problem was that the forest cadastre data is only renewed once every 10 years and it sometimes produces discrepancies with the IACS data. Also it should be mentioned that the results from this indicator should not be considered in isolation before concluding whether the measure had a positive or a negative effect. Other parameters, like habitat connectivity, habitat patching, should also be considered and the evaluation should only be concluded from consolidated results.

4. Strengths and weaknesses of the indicator

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Easy to make comparison, because the new ecotones are easy to monitor not depending on the scale	For quality impact (this indicator only shows the changes of ecotone in length, the quality is not assessed) assessment, there is a need to collect data on site because available datasets are inadequate for such assessment	Spatial changes in landscape and HNV territory are easily distinguished
Timing of environmental impacts captured	The spatial change of the ecotone is captured after the measure is implemented	For quality assessment at least 5 years permanent monitoring is needed to collect additional data	Spatial changes can be captured every year because IACS-LPIS georeferenced data are available every year
Establishment of robust causal relationships	The calculation of the indicator shows direct influence in the quantity of ecotone	To be able to assess impact on quality additional research and data are needed	The application of measures show direct influence on landscape heterogeneity changes
Establishment of consistent micro-macro linkages	Methodological approach explicitly covers and combines micro and macro level analysis. Consistency and validation procedures are internalised.	None	Macro-level analysis can be built on aggregated micro-level results. This approach establishes linkages with variables that are suitable to upscaling.
Assessment of net-impacts	None	This method is not a stand-alone method. To assess net-impacts, other indicators have to be taken into account.	Before and after analysis is limited only to the changes observed.

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Appropriateness of indicator(s) to capture complexity of environmental relationships	It is easy to capture one aspect – RDP impact on heterogeneity of landscapes	It is not enough to use the proposed indicator to be able to measure complexity of environmental relationships	Method relies on spatial data on application of measures. Ground truth survey data can be used to address data gaps.
Unambiguous and understandable results	Results are easy to understand and communicate, no specific technical skills are required	To present complexity of environmental relationships more results from other indicators are needed as this indicator as single is not sufficient	Method provides user friendly outcomes.

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

References

Fagan WF, Fortin MJ, Soykan C (2003) Integrating edge detection and dynamic modeling in quantitative analyses of ecological boundaries. *Bioscience* 53:730-738.

Hansen AJ, di Castri F, Naiman R (1988) Ecotones: what and why? *Biology International* special issue 17 [A new look at ecotones: Emerging international projects on landscape boundaries]: 9-46.

Wiens JA (1992) What is landscape ecology, really? *Landscape Ecology* 7:149-150.

Indicator: High Nature Value farmland

1. Definition / description of the indicator, including:
   - *Environmental public good*: Biodiversity HNV farmland
   - *Type of indicator*: High Nature Value (HNV) farmland is typically characterised by a combination of low intensity land use, the presence of semi-natural and unfarmed features and a diversity of land cover and land uses, supporting the presence of high-level biodiversity of wildlife species and habitats. HNV farmland and HNV farming systems are composite indicators. The basic components of these indicators are represented by: 1) high proportion of semi-natural vegetation; 2) mosaic of low-intensity agriculture; 3) supporting wild species and habitat of conservation concern.
   - *Reflected RDP priority and focus area*:
     - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
     - Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.
   - *Unit of measurement*: a) Percentage of HNV farmland on UAA and b) HNV score at farm level
   - *Type of data required*: In case of measurement a): georeferenced data on land cover and land use with sufficient details to guarantee the assessment of semi-natural features, the level of farming intensity and the presence of wildlife species and the possibility of comparison between participants and non-participants (e.g. LPIS database). In case of measurement b): individual data of samples of farms with information on crops, livestock and type of farming practices.
   - *Scale and level of application*: HNV farmland may exist at different scales from single parcel to an entire landscape, while HNV farming system refers to land cover and associated farming practices of the system as a whole, either it is at farm level or at landscape level.

2. Existing data sources
   - *EU-level*: FADN, IACS, LPIS, JRC maps on HNV and semi-natural vegetation
   - *MS and regional level* (examples): Data on land cover, farming intensity (nitrogen and pesticide), and ecological quality index are available through the Regional Environment Agency (ARPAV) and Managing Authority (Veneto Region). Farmland Bird Index data from National Rural Network
   - *Fact sheets and information available from other sources*: Paracchini et. al, 2009; EENRD 2010; Keenleyside et al. 2014.

3. Context of the case study testing
   - *Case study area*: Veneto Region - Italy.
   - *Policy context*: Natural handicap payments to farmers in mountain areas (211) Agri-environment measures aimed to increase biodiversity 214/A (Ecological corridors, buffer

strips, hedgerows and thickets), 214/C (Organic farming); 214/D (Protection semi-natural habitats and biodiversity), 214/E (Meadows and grasslands); 214/F (Biodiversity) and Support for non-productive investments (216)

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

4. Strengths and weaknesses of the indicator

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

5. Recommended application

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

availability of data concerning large and small patches of perennial vegetation detected in fine-resolution satellite images should increase the reliability of land cover in agro-ecosystems at reasonable monitoring costs.

References

EENRD (2008) Guidance Document to the Member States on the Application of the HNV Impact Indicator. European Evaluation Network for Rural Development, Brussels.

García-Feced C. Weissteiner CJ. Baraldi A. Paracchini M.L. Maes J. Zulian G. Kempen M. Elbersen B. Pérez-Soba M. (2015) Semi-natural vegetation in agricultural land: European map and links to ecosystem service supply, Agronomy for Sustainable Development, Vol. 35, Issue 1, pp 273-283.

Keenleyside, C, Beaufoy, G, Tucker, G, and Jones, G (2014) High Nature Value farming throughout EU-27 and its financial support under the CAP. Report Prepared for DG Environment, Contract No ENV B.1/ETU/2012/0035, Institute for European Environmental Policy, London.

Paracchini M-L, Petersen J-E, Hoogeveen Y, Bamps C, Burfield I, van Swaay C et al. (2009) “Identification of High Nature Value Farmland at the EU27 Level on the Basis of Land Cover and Biodiversity Data”, Joint Research Centre, Ispra (IT).

Biodiversity Wildlife

Indicator: Number of farmland bird individuals

1. Definition / description of the indicator, including:
    - *Environmental public good*: Biodiversity Wildlife
    - *Type of indicator*: Additional programme-specific indicator
    - *Reflected RDP priority and focus area*:
        - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
        - Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.
    - *Unit of measurement*: Number of farmland bird individuals
    - *Type of data required*: Regularly collected biodiversity related (bird census) data from previously set geological location in a timescale of the programme period under consideration
    - *Scale and level of application*: Biodiversity related data is evaluated at the level of the survey points of the observation

2. Existing data sources
    - *EU-level*: Data for the number of farmland bird individuals is the baseline data for Farmland Bird Index. Data collection (monitoring) standards of the common bird species is set by the European Bird Census Council. Results of the Farmland Bird Index estimations at EU level are available at the following website: http://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_population_trends_of_farmland_birds
    - *MS and regional level* (examples): In Hungary the Common Bird Monitoring Programme has been running since 1999. A database of approximately 300 2.5x2.5 km survey squares is available for the whole timescale at the Monitoring Center of BirdLife Hungary. Detailed descriptions can be found at the following website (in Hungarian): http://www.mme.hu/mindennapi-madaraink-monitoringja-mmm
    - *Fact sheets and information available from other sources*: The proposed indicator is an alternative application of the Farmland Bird Index data sources; no direct fact sheet is available.

3. Context of the case study testing
    - *Case study area*: Heves Plain High Nature Value Area , Hungary
    - *Policy context*: Biodiversity decline is well-known throughout Europe, with agricultural habitats facing significant challenges. Pillar 2 measures can contribute to halt this overall decline. As several scientific studies and the programme evaluations show, well-targeted agri-environmental measures may hinder the further decline in agricultural biodiversity.

Heves Plain High Nature Value Area is one of the most successful HNV area in terms of the uptake of the AE measures, thus provides a good opportunity for comparing the biodiversity values of participant and non-participant survey points. As the landscape is scattered by mosaic-like natural habitats (grasslands, wetlands, etc.) during the case study testing naturalness of the areas was also taken into consideration.

- *Used data*:
  1. Biodiversity data for 19 survey squares have been used for the last programming period of agri-environmental measures (2009-2014). In each monitoring square, data for 15 survey points are available, representing micro-level data for the exact location and within a 100m radius of the survey.
  2. Participation data of the agri-environmental measures were used based on the Land Parcel Identification System.
  3. Naturalness of the areas under examination was assessed by using CORINE 1:50 000 land cover data base.
- *Evaluation approach tested*: A number of farmland bird individuals of the 285 survey points were compared based on the detailed grouping of the available data sets. Group design was based on the AE measure participation and the 'naturalness' of the survey points (share of the participant area inside the survey point/share of natural areas inside the survey point). Participant-natural, participant-non-natural, non-participant – natural, non-participant – non-natural groups were created, where the number of farmland bird individuals was assessed in parallel at the above mentioned time scale. Group design was carried by using spatial analyses tools (Jenks Natural Breaks method).

4. Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	Survey spots representing micro level by observing biodiversity data at parcel level.	Less frequently updated land-use data can limit the application of the indicator in elaborate statistics-based evaluations and result in unobserved impacts at local level	By using baseline data of a widely-known indicator, the proposed indicator may contribute to easier analyses of micro-level RD measure impacts.
Appropriateness of indicator(s) to capture complexity of environmental relationships	As the biodiversity data used is available as the baseline data of the Farmland Bird Index, data quantity limitations are not expected. Large sample size enhances the possibility of using multiple comparison groups as well as elaborate statistics-based methods to filter out other intervening factors.	Number of farmland bird individuals as an overall biodiversity indicator shall be further developed as this is rather sensitive to the effects of different years (weather conditions, migration circumstances, etc.).	As the data sources used are available in most of the EU Member States, the approach has high potential in replicability. More robust counterfactual assessment at micro level possible compared to using the FBI.

5. Recommended application
   - The use of the indicator of ‘Number of farmland bird individuals’ is recommended in cases where micro-level impacts of the different RD measures shall be detected, but biodiversity data gaps are observed at parcel level and the FBI cannot be used.
   - Baseline data of Common Birds Monitoring Programme shall be available, which means that the cooperation with the relevant monitoring organisations is highly recommended.

Indicator: Number of singing corncrake males

1. Definition / description of the indicator, including:
   - *Environmental public good*: Biodiversity Wildlife
   - *Type of indicator*: Additional (measure specific) indicator
   - *Reflected RDP priority and focus area*:
     - Priority 4 of the RD programmes: Restoring, preserving and enhancing ecosystems related to agriculture and forestry.
     - Focus area 4A: Restoring and preserving biodiversity, including in Natura 2000 areas and high nature value farming, and the state of European landscapes.
   - *Unit of measurement*: Number of singing corncrake males
   - *Type of data required*: Regularly collected data on singing males of corncrakes, land cover data and agricultural land-use data (IACS).
   - *Scale and level of application*: The indicator is tested at micro level, scale of sampling plot – 0.28 km^2 (observation radius – 300 m).

2. Existing data sources
   - *EU level*: corncrake singing males census data is not systemically gathered, but is available in the countries which report about conservation status of corncrake according to the reporting requirements for EU Birds Directive implementation. Other necessary data, such as land-cover data (CORINE land cover or national equivalents), agricultural land-use data (IACS) is typically available.
   - *MS and regional level* (examples): In Lithuania corncrake census data is gathered within the framework of the state biodiversity monitoring programme. The monitoring is performed in the Natura 2000 areas designated for the conservation of this species.
   - *Fact sheets available from other sources*: no direct fact sheet is available.

3. Context of the case study testing
   - *Case study area*: Nemunas delta regional park, Šilutė municipality district, Lithuania.
   - *Policy context*: The indicator focuses on the impact of RDP Measure 214, which is one of the key measures to address biodiversity decline in grasslands under the CAP policy. In addition, the evaluation context directly relates to the EU biodiversity strategy implementation, in particular, target 3 “increase contribution of agriculture and forestry to biodiversity”.
   - *Used data*:
     1. Corncrake density data are used collected from the national state biodiversity monitoring programme. There were 115 observation sample plots (circular form, radius 300 m) included in the evaluation.
     2. Georeferenced spatial data set at a scale of 1:10,000 in the Republic of Lithuania (GDR10LT), Orthophoto images 2010-2014
     3. Integrated Administration and Control System (IACS) available in GIS format for period of 2010-2013.

4. Additional databases were used: forest cadastre databases, Corine Land Cover 1:50,000. These databases were used for species environment analysis to determine side effects.

- *Evaluation approach tested: testing focussed on defining causal linkages between the* occurrence of corncrakes and RDP measures, using data generated by the biodiversity monitoring programme carried out on a regular basis by the public sector. This involved taking evaluation steps through the developed logic model and analysing evaluation results eliminating various factors possibly impacting on evaluation conclusion. Functional unit for the corncrake was selected also on circular shape (diameter 600 m, covering 0.28 km^2) corresponding with standard observation point area as defined in corncrake monitoring methodology. Functional units were grouped according to the participation rate of the evaluated AEM within the sample areas. The functional units were grouped into multiple comparison groups according to intensity of participation in the targeted measure. The counterfactual scenario was defined based on with and without involving observation plots without participation of targeted measure, but with similar natural conditions. Multiple regression analysis was applied to analyse dependency of the corncrake numbers in the observation plot with different participation intensity. Results of the case study illustrate a robust statistical relation ($p=0.01$) between the number of observed birds and the participation intensity of Measure 214 in the functional unit. More corncrakes were observed in the fields where participation rate is higher. Different density rates cannot be explained solely by mowing activities, as monitoring was done in June when some of the non-participants had not yet mowed their areas. The higher density can be partly explained by the different (more natural) structure of the vegetation caused by participation in Measure 214, which leads to higher availability of food. However, if the monitoring would be performed later (or even better, a third count enabling comparison of the changes), results would be more informative indicating a clearer discrepancy between participants and non-participants.

4. Strengths and weaknesses of the indicator (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Compatibility with local environmental and farm structural characteristics	The indicator can deliver robust data at micro level considering specific local characteristics of the assessed areas, which allows the evaluation at micro level	The indicator can be performed only for specific sub-measures within Measure 214 linked to limited types of grassland.	The indicator would contribute to more comprehensive evaluation as additional indicator together with FBI evaluation.
Appropriateness of indicator(s) to capture complexity of environmental relationships	Reflects robust causal relationships of biodiversity impacts of specifically targeted sub-measures of AEMs	The indicator is linked to narrow aspect of environmental problem/public good. The indicator shall be used in combination with other indicators e.g. FBI	The indicator provides an example for additional result indicators needed to assess biodiversity

Evaluation challenges (relevant for indicators)	Strengths	Weaknesses	Key contribution to evaluation benefits
		For a wider application of the indicator and an application with robust statistical counterfactuals adjustments to the corncrake census data methodology are needed (e.g. set later timing for the second count of birds).	impacts of specifically targeted sub-measures of AEMs.

5. Recommended application
   - Number of singing corncrake males is recommended for the evaluation of impacts of Measure 214 at micro level. It could be relevant to use this indicator as an additional one along with FBI evaluation. Such an approach would contribute to FBI by providing additional information at micro level, while FBI itself is more a macro-level indicator.
   - More adjustments are needed in adapting the data gathering methodology for RDP evaluation usage. In addition, more research is needed to evaluate corncrake breeding success (and determine best timing for mowing) for this species to be an 'umbrella indicator'.
   - The indicator provides a good example of collaboration and data sharing potential between Agriculture and Environment sectors. There might be other data gathered by environmental sector, which could be successfully used for environmental evaluations of RDP measures.
   - It would be beneficial to have more coordination between environmental and agriculture authorities to determine data-sharing mechanisms. With adjustment, data could be gathered within the framework of the existing state biodiversity monitoring programme, leading to no additional costs (or a comparatively insignificant increase due to e.g. additional counts).
   - At the moment, such data is gathered only within designated Natura 2000 sites. It would be relevant to select a statistically robust number of samples outside protected areas, which would enable modelling results on macro level and national scale.

References

Berg A, Gustafson T (2007) Meadow management and occurrence of corncrake Crex crex, Iris pseudacorus, Phragmites australis and Phalaris arundinacea. *Agriculture, Ecosystems and Environment* 120: 139–144

Budka M, Osiejuk T (2013) Habitat preferences of Corncrake (Crex crex) males in agricultural meadows. *Agriculture, Ecosystems and Environment* 171: 33–38.

Green RE (1996) Factors affecting the population density of the corncrake Crex crex in Britain and Ireland. Royal Society for the Protection of Birds, 17 Regent Terrace, Edinburgh, EH7 5BN, UK

Keišs O (2005) Impact of changes in agricultural land use on the Corncrake *Crex crex* population in Latvia. *Acta Universitatis Latviensis* 691: 93–109.

Animal welfare

Indicator: Animal-based / result-based indicators: Lameness and mortality rates

1. Definition / description of the indicator, including:
   - *Public good*: Animal welfare, Animal welfare category: Good health
   - *Type of indicator*: Programme-specific result indicators. No common result or impact indicators exist and managing authorities and evaluators are not formally required to define additional result or impact indicators targeted at animal welfare. However, for programmes which have implemented measures targeted at animal welfare (e.g. Measure 215 and 121), it is necessary to define and select suitable indicators to assess the effects of those measures on animal welfare. The sole use of output indicators is not sufficient. Animal-based indicators integrate a direct and result-based approach into the evaluation of animal welfare impacts. Lameness and mortality rates of cows and calves are two of the direct, i.e. animal-based, indicators to measure changes in the animal welfare category 'good health', established in the Welfare Quality® protocol. The indicator lameness measures changes in the share of lame animals compared to the total number of animals, while the indicator mortality rates measures the share of dead animals. The indicator lameness has causal linkages with policy measures targeted at improving housing conditions, such as type of bedding and the provision of straw, the provision of access to grazing and improving health care plans. The indicator mortality rates has causal linkages with policy measures targeted at improved feeding and water access, improving housing conditions such as type of beddings and space allowances and improving health care plans.
   - *Reflected RDP priority and focus area*: No focus area is particularly defined in relation to animal welfare in the CMES, but animal welfare is included in the rural development priority 3 "Promoting food chain organisation, including processing and marketing of agricultural products, animal welfare and risk management in agriculture".
   - *Unit of measurements*: Share of lame animals and share of died animals
   - *Type of data required*:
     - Type of data required to measure / quantify the indicator:
       - Animal-based data: Livestock monitoring data either available from secondary data sources (e.g. the HIT database in Germany) or from empirical monitoring efforts through farm visits
     - To enable the application of the indicators in an assessment of net-effects with advanced evaluation methods additional type of data are required:
       - Livestock husbandry and farm structural data: Data on husbandry systems and farm structural characteristics available from secondary data sources (e.g. FADN, Census data etc.)
       - Policy related data: IACS data on uptake of relevant measures (IACS database).
   - *Scale and level of application*: Farm level

2. Existing data sources
   - *EU-level*: No EU-wide data sources exist.
   - MS and regional level (examples): The indicators are collected as part of the benchmarking system of the Animal Health and Welfare Management Programme in Scotland (Measure 215) and the indicator mortality rates is included in the HIT database in Germany.
   - *Fact sheets and information available from other sources*: No particular fact sheets exist, but Welfare Quality® Assessment Protocols have been developed by the Welfare Quality® consortium (2009).

3. Context of the case study testing
   - *Case study area*: North Rhine Westphalia, Germany
   - *Policy context*: Animal welfare payments - Measure 215, 2. Investment support with animal welfare related objectives – Measure 121.
   - *Used data*: Empirical monitoring data from farm visits (winter 2013 / 2014), HIT database.
   - *Approach applied to review the suitability of the indicator* (short explanation of the main logic model steps): The testing focussed on the development of guidelines for the selection of animal welfare indicators for RDP evaluation covering different relevant animal welfare criteria. Following the identification of the most relevant (and practical) animal welfare criteria which need and can be covered by the evaluation, different types of animal welfare indicators were reviewed and tested to inform the development of the guidelines for indicator selection. The case study differentiated between indirect indicators such as in relation to management and housing, and direct indicators such as in relation to animal health. Based on a review of stakeholder acceptance and practical feasibility of direct animal-based indicators (Bergschmidt et al., 2014 and 2015) advantages and disadvantages of different indicator types and the strengths and weaknesses of the integration of different result-based indicators in the animal welfare assessment were derived.

4. Strengths and weaknesses of animal-based indicators (based on SWOT)

Evaluation challenges (relevant for indicators)	**Strengths**	**Weaknesses**	**Key contribution to evaluation benefits**
Appropriateness of indicator(s) to capture complexity of animal welfare relationships [Lack of suitable animal welfare indicators in RDP evaluations]	Adds a direct (i.e. result-based) assessment of health criteria to the assessment of housing and feeding (water) criteria through the use of resource or management based indicators High acceptance by stakeholders and scientists Cost-effective application in combination with resource and management-based indicators feasible	Single indicator limited health aspects Cost-effective application depends on available monitoring data. High monitoring requirements and costs might prohibit the application if no data sources exist. Indicator can be influenced by seasonality	Improves the coverage of animal welfare impacts and contributes to a conceptually sound multi-criteria assessment of animal welfare

5. Recommended application
    - The application of the indicator is lameness recommended for the evaluation of animal welfare payments and investment support with an intervention logic linked to health and housing animal welfare criteria (micro level). The indicator mortality rates of cows and calves is best used in a multi-criteria assessment in combination with indicators on grazing access, increased space allowance and walking surface (resource and management-based indicators). The application of the indicator is recommended for the evaluation of animal welfare payments and investment support with an intervention logic linked to feeding (water), health and housing animal welfare criteria (micro level). The indicator is best used in a multi-criteria assessment in combination with indicators on feeding and water access, type of beddings and space allowances, and walking surface (resource and management based indicators).
    - The farm visits and livestock monitoring conducted in a pilot project by the Thünen Institute (Bergschmidt et al., 2015) highlight the high amount of staff resources required to monitor a sufficiently large sample in different years for RDP evaluations of animal welfare impacts. This implies that the feasibility of using these indicators in RDP evaluations depends on the availability of already existing monitoring data or secondary data sources. In case of long-term evaluation contracts, different sampling strategies can be explored to collect primary data through farm visits.
    - Practitioners and farmers viewed had concerns about the use of the indicator mortality rates, as they felt that on small farms the occurrence of one accident or disease could already affect their eligibility for payment. This problem can however be solved by using average mortality rates over several (e.g. three) years.

References

Bergschmidt A, Renziehausen C, Brinkmann J, March S (2014) Tiergerechtheit landwirtschaftlicher Nutztierhaltung: Verbesserung durch ergebnisorientierte Honorierung? *Ländl Raum* (ASG) 65(2): 32-33.

Bergschmidt A, Renziehausen C, March S, Brinkmann J (2015) Tierschutzwirkungen der Entwicklungsprogramme für den ländlichen Raum - Ergebnisse der Evaluierung der Maßnahme "Förderung umwelt- und tiergerechter Haltungsverfahren". In: KTBL, Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V. (Hrsg.): Herausforderung Tierwohl, KTBL-Tage: 208-223.

Welfare Quality® consortium (2009) Welfare Quality® Assessment Protocol for Cattle ISBN/EAN 978-90-78240-04-4, 180 pages.