ENVIEVAL

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

(Project Reference: 312071)

Area 2.1.4: Socioeconomic research and support to policies

KBBE.2012.1.4-08: Development and application of methodologies and tools, including indicators, for the assessment of environmental impacts of rural development programmes in the EU

Report D5.1

Review of macro level methods and scales

Authors: Inge Aalders, Jane Morrice, David Miller, Rupert Hough (JHI), Andrea Povellato, Davide Longhitano (INEA), Gerald Schwarz (TI)

Approved by Work Package Manager of WP5: David Miller (JHI)

Date: July 2013

Approved by Project Coordinator: Gerald Schwarz (TI)

Date: July 2013

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

SEVENTH FRAMEWORK PROGRAMME

Table of Contents

List of Figures

List of Tables

List of Acronyms

AES	Agri-Environmental Scheme
AFI	Agri-environmental Footprint Index
ATC	Ability To Choose
BBS	Breeding Bird Survey
CA	Conjoint Analysis
CAP	Common Agricultural Policy
CBA	Cost Benefit Analysis
CMEF	Common Monitoring and Evaluation Framework
CSS	Countryside Stewardship Scheme
EF	Ecological Footprint
ES	Environmental Stewardship
ESA	Environmentally Sensitive Area
FADN	Farm Accountancy Data Network
GAM	Generalized Additive Model
GHG	GreenHouse Gas
GLM	General Linear Model
GMM	Generalized Mixed effect Model
GTAP	Global Trade Analysis Project
HARM	HARMonized regions
HNV	High Nature Value
IACS	Integrated Administration and Control System
LCA	Life Cycle Analysis
MARS	Monitoring Agriculture with Remote Sensing
MCA	Multi Criteria Analysis

MCMC	Markov Chain Monte Carlo
MDS	Multi-Dimensional Scaling
NUTS	Nomenclature of Territorial Units for Statistics
OELF	Ecologically, Culturally and Politically Significant Area
RDP	Rural Development Programme
SIMPER	SIMilarity PERcentage
SPARD	Spatial Analysis of Rural Development
SRG	Stakeholder Reference Group

Executive Summary

The Common Monitoring and Evaluation Framework (CMEF) is used for the evaluation of Rural Development Programmes (RPD) (European Commission, 2006). The key methodological challenges for the environmental impact evaluation in relation to public goods are: to provide evidence of true causality, to disentangle the effects of single measures and the programme from other factors, to quantifiy net-impacts at macro level, the availability of viable body of evidence and closing the gap between indicator measurement and policy decision-making.

The objective of ENVIEVAL is to identify and test suitable methods for the assessment of net-impacts of RD measures and programmes against a changing baseline of pressures. These assessments need to be carried out at a scale appropriate to the representation of natural processes In addition, the methodological framework to be developed in ENVIEVAL needs to support the analysis of multiple benefits at the most relevant scale and to include the potential for cumulative environmental impacts. A comprehensive assessment of the extent to which the multiple environmental goals for the RDP have been achieved requires more than a single methodology. This review explores a range of methodologies which may address some of the challenges and contribute to the development of a flexible integrated methodological framework for the assessment of environmental impacts of RDPs at macro level. The main objectives of the review are:

- To review and define suitable scales to consider as 'macro level'
- To review the application of macro-level methodologies in previous and current evaluations
- To review new methodological developments which address existing challenges in macro-level evaluations
- To recommend candidate macro-level methods for testing in the public good case studies.

The methodologies for a macro-level assessment of environmental impact of the RDP need to overcome a number of challenges including an integration of scales (RDP and public goods), a selection of an appropriate scale for measurements/indicators, the establishment of causality between overall RDP activity and changes in public goods, as well as assessing environmental impact at and across different geographical levels. Understanding the issue of scale in the environmental impact assessment of RDP is critical to resolve the current challenges for the macro-level assessment.

Landscape science and hierarchy theory provide an important theoretical and methodological basis to develop this understanding. The definition of landscape by the European Landscape

Convention is "An area as perceived by people, whose character is the result of the action and interaction or natural and/or human factors." Based on this definition 'landscape scale' refers to a spatial scale above the field-, farm- and local scale; and can be a catchment, an area of coherent landscape character or a sub-unit of a natural region.

Spatio-temporal hierarchies in landscape processes allow classification according to temporal and spatial scale/dimension. The advances in relation to understanding complex systems in this way are based on the dynamic processes in those ecosystems. In order to assess the impact of RDP measures on the environment, it is important that both the action and the measurement are placed in the appropriate context of the complex system. Hierarchy theory may suggest that the measurement/indicator should be at a higher level along the scale than the RDP action.

The review of recent RDP evaluations in Austria, Finland, France, Germany, Great Britain (England and Scotland), Greece, Hungary, Italy, Lithuania and the Netherlands included the reported use of macro level (i.e. programme-level) assessment, which is reported here.

Priority was given to review reports in the following order:

1. 2007-2013 On-going evaluation reports and available ex-post evaluation results, cases with macro level reviewed
2. 2007-2013 Mid-term evaluation, cases with macro level reviewed
3. 2007-2013 Strategic Environmental Assessment, cases with macro level reviewed
4. 2000-2006 Ex-post evaluation reports, x cases with macro level reviewed
5. 2000-2006 Mid-term evaluation reports, x cases with a macro level reviewed
6. Other relevant RDP evaluation reports, x cases with a counterfactual reviewed.

We have collated the reported macro-level (programme) assesments explicitly mentioned in the evaluation reports. Thus the accuracy and specificity of the original evaluation reporting drives our review results. For identified macro-level assessment we report on the type of environmental impact indicator, type of analysis, scale, causality, net impact and link between micro and macro. Where possible the type of method used for analysis is classified broadly into one of the five categories: multi-scale sampling, integrated/landscape, quantitative and statistical analyses, scaling and qualitative approaches.

In general the results of the review of macro-level methodologies in the evaluation reports are patchy. In many cases the macro-level evaluation is based on qualitative assessment often using the evaluation questions (CMEF) or expert/evaluator assessment, which suggests an absence of evidence for causality between RDP and delivery of public goods. The reported quantitative or analytical methods are almost as common as the qualitative methods. These methods seem to be simple tallies in many cases and do not explicitly and clearly include a disentanglement of single measures and overall programme effects from other factors. However there are a number examples of more complex and integrated methodologies and multiscale approaches that do aim to address this issue (Lower Saxony – HNV and water quality, Austria - wildlife, Veneto – landscape, and France – water quality).

The methodologies vary per public good. The reported methods include:

- Climate change – estimation and aggregation
- Water quality – impact coefficient, index for level of pollution of macrodescriptors (LIM), OECD methods, monitoring networks, and case studies
- Biodiversity wildlife – paired comparisons between areas participating and non-participaring
- Soil – modelling erosion loss
- Landscape – Shannon index and choice experiments
- Animal – disaggregate analysis

Overall, there is a demand for more integrated methodologies as reported by Austria and Lower Saxony which also enable the assessment of net environmental impacts at macro level. The review of the recent evaluation reports confirmed the key challenges for macro-level evaluations highlighted in the introduction and emphasised the need for the integration of new methodological developments.

The emphasis of the review of new methodological developments is specifically on methods for a macro-level assessment of the environmental impact of the RDP and its key challenges: determination of true causation, aggregation of all economic, social and environmental net-impacts of RDP, and a cost-benefit analysis. Therefore, a modified classification of approaches has been chosen for the methodologies to reflect the demand for methods that can address complex challenges: statistical approach, hierarchical and multi-scale approach, spatial analytical approach, multi-criteria approach, and integrated/landscape approach.

The results from the review show clearly that a lot of methodological progress has been made in recent years to improve a range of aspects regarding the environmental impact assessment of RDP measures, although many methods to date are only applied to ex-ante rather than ex-post assessment. The most important methodological developments are the advances made in relation to multi-criteria, spatial analytical approaches and integrated approaches, as well as efforts made to address the scale mismatch between economic and ecological/natural sciences. These developments are able to contribute to addressing the challenges posed by the demand for measuring the impact of RDP activities/investment on the delivery of public goods.

There is evidence that just tallying the implementation of the measures does not equate to successful delivery of public goods and that it does matter to the success of RDP where and how measures are implemented. At the same time the issue of scale in relation to measuring impact has also been shown to be important. Evidence of cause and effect between RDP and public goods is largely limited to biodiversity (wildlife) and water quality.

The methodologies used for the macro (programme level) assessment is patchy and in many cases is not actually making an assessment at programme level but at measure level. Multi-criteria analysis, in particular agri-environmental footprinting, is the most promising approach to assess the complexity of assessing multiple measures. However a key challenge will be to take this meaningfully from farm level to regional and national level. In this context, lessons from other methodological approaches for scaling and conceptual integration may be valuable.

Based on the developments towards integrated assessments, macro-level assessment should probably focus not just on the overall programme assessment (i.e. effectiveness of investment), but more to the wider environmental impact which can be at different geographic scales. This does require the inclusion of spatially-explicit analysis at appropriate scale levels. Previous EU projects have addressed a number of relevant issues in the effort for integrated analysis.

Recent developments in spatial econometrics in principal address the need to include spatially-explicit analysis at appropriate scale levels. The case study applications of new spatial econometric models in the SPARD project have indicated that such models have the potential to assess and quantify net-impacts at macro level, which was confirmed as one of the key challenge for macro-level evaluations in the stakeholder interviews and workshop discussions. Including spatial econometric models in the ENVIEVAL case studies would enable the testing of the suitability of spatial econometric models to address the challenge to assess net environmental impacts of rural development measures and programmes at macro level. However, a potential limitation for the use of spatial econometric models is their large and complex data requirements. The data requirements and availability need to be assessed in more detail in the next task to ensure that feasible case study areas with sufficient data infrastructure can be selected and validated with the SRG.

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

The main gaps in the current knowledge and candidate methods for the case study testing are synthesised below.

Main gaps in the current knowledge:

- limited evidence of causality between the RDP objectives and indicators
- limited evidence of the ability of indicators to measure impact across and within scales and levels
- the need to incorporate the spatial context around participating areas.

Candidate methodologies:

- Increasing evidence base
    - Hierarchical approaches
        - systematic and consistent sampling method (Gabriel et al., 2006, 2010; Concepcion et al., 2012; Deconchant et al., 2007)
        - use of typologies (Righi et al., 2011; Andersen et al., 2007; Teillard et al., 2012)

        - scaling (up- and downscaling - Ewert et al., 2011)
    - Spatial analytical approaches –
        - landscape metrics - the flexible whole-landscape modelling framework by Ferrier and Drielsma (2010)
        - spatial econometrics – by Reinhard et al., 2013; Reinhard and Linderhof, 2013
- Assessment
    - Multi-criteria analysis – agri-ecological footprint
    - Spatial econometrics
    - Landscape zoning and multi-functional hotspots

1 Background

1.1 Introduction

The Common Monitoring and Evaluation Framework (CMEF) is used for the evaluation of Rural Development Programmes (RPD) (European Commission, 2006). The key methodological challenges for the environmental impact evaluation in relation to public goods are: to provide evidence of true causality, to disentangle the effects of single measures and the programme from other factors, to quantifiy net-impacts at macro level, the availability of viable body of evidence and closing the gap between indicator measurement and policy decision making (Lukesch & Schuh, 2010). In addition, across Europe there are marked differences in the implementation of the RDP itself and the indicators across the EU (Dwyer et al., 2007; Zucker, 2006) and heterogeneity of rural environments and of the public goods provided by agriculture. Measures are often implemented in combination which make it difficult to disentangle the impacts of single measures (Cooper et al., 2006) and many environmental impacts of RDP measures are site-specfic (Stolze et al., 2000; Whittingham, 2011). This has raised the demand for a flexible evaluation framework (Mortimer et al., 2010) and integrated methodological development (Metis GmbH and AEIDL, 2008).

The CMEF uses a systematic approach based on intervention logic which links a hierarchy of policy objectives to a hierarchy of indicators which aim to measure the extent to which the objectives are met. However, the intervention logic is not yet supported by robust empirical evidence of causality; in many cases it is based on common sense (Primdahl et al., 2010) and questions have been raised whether it is appropriate and possible to assess the environmental impact of RDP effectively in that way. Kleijn et al. (2001) were the first to question the effectiveness of AES on biodiversity. In the context of EASY, Kleijn and Sutherland (2003) reviewed the effectiveness of agri-environment schemes on biodiversity. Based on 62 studies they concluded that there is limited research to support an assessment of the effectiveness of agri-environmental schemes. CMEF has been criticized by the European Court of Auditors (2011) for the lack of verifiable quantitative indicators resulting in a reliance on output indicators and the lack of cost-effectiveness assessment of indicators, monitoring systems and evaluation methods.

The environmental impact of RPD is assessed at programme level through only a small number of impact indicators: reversing biodiversity decline, maintenance of high nature value farming and forestry areas, improving water quality and contribution to combating climate change (European Commission, 2006). In recent years progress has been made in relation to the challenges of RDP evaluation, including an integrated ex-ante assessment of agri-environmental policies (SEAMLESS), interdisciplinary modelling and quantitative methods (ADVANCED-EVAL), and integrated farm-level assessment (AE Footprint). The SEAMLESS tool is a comprehensive method of integrating the multiple aspects of agricultural systems which includes aspects of the wider environment only as a resource or driver for the agricultural system; SEAMLESS does not incorporate the delivery of public goods through agricultural management. ADVANCED-EVAL has developed advanced (econometric) quantitative methods to improve ex-ante and ex-poste RDP evaluations. AE Footprint has developed an integrated farm-level assessment of RDP. An assessment framework was developed based on existing AE schemes for three broad policy objectives on agri-environment issues (natural resources, biodiversity and landscape) and three broad aspects of farm management (crop and animal husbandry, physical farm infrastructure, and natural & cultural heritage).

The objective of ENVIEVAL is to identify and test suitable methods for the assessment of net-impacts of RD measures and programmes against a changing baseline of pressures. These assessments need to be carried out at a scale appropriate to the representation of natural processes. In addition, the methodological framework to be developed in ENVIEVAL needs to support the analysis of multiple benefits at the most relevant scale and to include the potential for cumulative environmental impacts. A comprehensive assessment of the extent to which the multiple environmental goals for the RDP have been achieved requires more than a single methodology. This review explores a range of methodologies which may address some of the challenges and contribute to the development of a flexible integrated methodological framework for the assessment of environmental impacts of RDPs at macro level. The main objectives of the review are:

- To review and define suitable scales to consider as ‘macro level’
- To review the application of macro-level methodologies in previous and current evaluations
- To review new methodological developments which address existing challenges in macro-level evaluations
- To recommend candidate macro-level methods for testing in the public good case studies.

2 The Issue of 'Scale'

The meaning of 'macro-level' in the context of environmental impacts of RDP is ambigous, because it depends on the scale used to measure the impact. In this section we explore the current discourse on scale that lies at the heart of the RDP's challenges for the assessment of environmental impact.

Both natural and social scientists use the concept of scale and report by spatial units. However they base their application on very different theoretical backgrounds. Consequently, the collection, analysis and reporting of data, as well as the analysis undertaken, can be significantly different (Gibson et al., 2000). Contributions to the discourse on this issue have emerged from both natural science and social science communities (Schneider, 2001, Keshkamat et al., 2012, Adger et al 2006), as well as from systems and complex science communities (Cash et al., 2006; Gibson et al., 2000; Veldkamp et al., 2011). The differences are best illustrated by Vermaat et al. (2005) in their study on the matching of scales in spatial economics and landscape ecology. It concludes that the evidence for the latter is collected from real-world landscapes and recognises a nested hierarchy in complex ecological systems (Veldkamp et al., 2011), while economists use data compiled by national bureaux of statistics, which are often aggregated to the level of administrative units (Vermaat et al., 2005). The Millennium Ecosystem Assessment (MEA, 2003) distinguishes the terms 'scale of observation' and 'scale of the phenomenon', in which the former 'scale' is a construct of human measurement and the latter is a scale that each natural phenomenon has. The scale is the vertical axis along which any objects of interest are ranked, like on a ruler, while level is a fixed rank or horizontal layer on a scale, where all units belong to the same category (Figure 1). Spatial units can be homogeneous spatial entities like pixels or landcover classes, heterogeneous spatial entities like administrative boundaries (counties, postal code areas) or temporal units (Fekete et al., 2010).

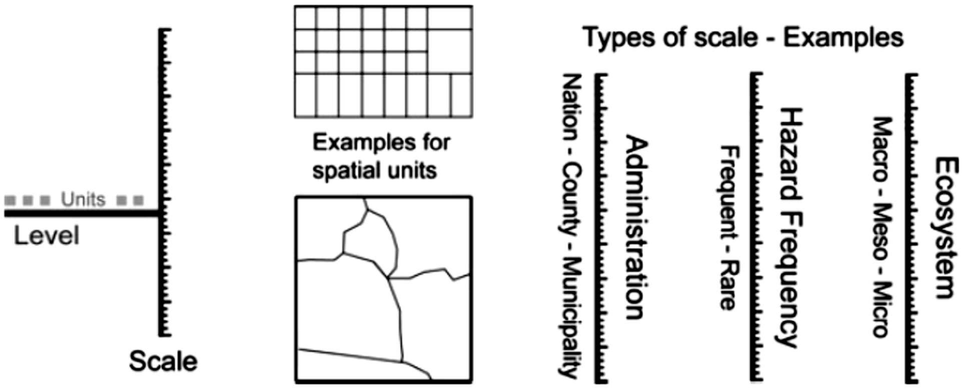

Figure 1 Explanation of the main working terminology used for scale, level and unit (Source: Fekete et al., 2010).

An extensive literature review into spatial scale mismatch by Pelosi et al. (2010) suggests that a solution for the spatial scale mismatch can be found in a systematic approach that integrates ecological and managerial processes and in a more accurate use of terminology and theoretical frameworks. Gibson et al. (2000) propose a definition of 'scale' as "*a spatial, temporal, quantitative, or analytical dimension used to measure and study a phenomenon* and 'level' as *a location along a scale as the unit of analysis that are located at different positions*".

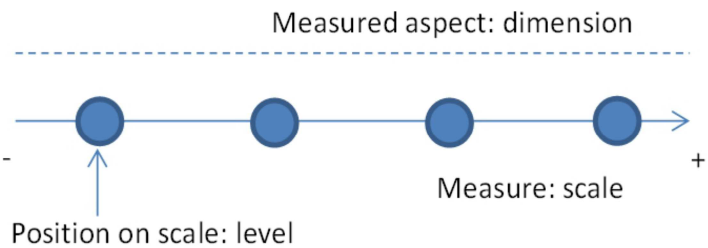

Figure 2 Schematic illustrations of scales, levels and dimensions (Source: Vervoot et al. (2011).

Vervoort et al. (2011) introduce the concept of 'dimension' (Figure 2) as the basic structure of analysis, which allows for recognition of multiplicity of possible scales. They identify levels (e.g. micro, meso, macro) as positions on a scale; however, when multiple scales are being used, the reference to macro or micro level can be ambiguous. The different possible scales that fall under this definition are commonly normative; an exception to this is cartography where the term 'scale' is a numerical ratio (e.g. 1:10,000) of the measured distance between two points on a map compared to the measured distance on the ground. Two maps that show the same two points on the ground, but on one map the distance is shorter than the other, then the scale of the former is smaller than that of the latter. This is referred to as a difference in resolution. So, for the same size of paper, it is possible to show a larger area of ground (this is the extent) with a smaller-scale map than a larger-scale map. As a guide, the national mapping agency of Great Britain (Ordnance Survey) refers to 'large scale' as 1:10,000 and greater, medium scale as 1:25,000 and 1:50,000, and smaller scale as anything smaller (e.g. 1:250,000). In Great Britain, field boundaries are mapped at scales of 1:25,000 and greater, which is the minimum scale used for mapping boundaries for fields for use in IACS. Fields of 2,000 ha and above must be shown on a map of at least 1:25,000, and those smaller than 2,000 ha on maps at a scale of 1:10,000 or greater. However, typically in environmentally-related studies, the term 'small scale' is used to refer to 'small area' relative to 'large scale' which is used to refer to 'large area'. There is no quantitative metric which can be used to clarify between small and large scale.

In Europe, a range of monitoring programmes (ecological, water, soil, landscape and land use) potentially can provide an important source of information to support the macro-level evaluation of RDP but this requires greater effort for data and methodological integration. There is a growing demand for integration which has led to research in integrated assessment in particular for the sustainability of agricultural systems. Integrated assessment recognises the following key challenges: aggregation versus disaggregation, treatment of uncertainty, integrating qualitative and quantitative knowledge, building up scientific and political credibility of integrated assessment models, and developing comprehensive and transparent scenarios (Bezlepkina et al., 2011).

Cash et al. (2006) discriminate different types of scales and levels, emphasising the cross-scale and cross-level interactions in managing the environment. As part of the scale review, we have identified a range of different scales that are relevant for RDP (Figure 3). The reference to macro level in the context of the RDP can refer to programme level, meaning the overall/aggregated achievements by measures in the four axes, as opposed to achievements by individual (sub)-measure or policy objectives. However, macro level can also refer to the wider context/public good (externalities) beyond the farm boundary (e.g. Concepcion et al., 2008). For the purpose of the review of methodologies, both the above interpretations of 'macro level' are considered.

The methodologies for a macro-level assessment of environmental impact of the RDP need to overcome a number of challenges including an integration of scales (RDP and public goods), a selection of an appropriate scale for measurements/indicators (Stoeglehner and Narodoslawsky, 2008), the establishment of causality between overall RDP activity and changes in public goods, as well as assessing environmental impact at and across different geographical levels (Lukesch and Schuh, 2010; Michalek, 2012). Understanding the issue of scale in the environmental impact assessment of RDP is critical to resolve the current challenges for the macro-level assessment.

Hierarchy theory and landscape science provide an important theoretical and methodological basis to develop this understanding. Spatio-temporal hierarchies in landscape processes allow classification according to temporal and spatial scale/dimension. Steinhardt and Volk (2003) examine a hierarchical nested approach of watershed modelling, drawing on hierarchy theory which "*suggests that when a phenomenon is studied at a particular hierarchical level, the mechanistic understanding comes from the next lower level, whereas the significance of that phenomenon can only be revealed at the next higher level*". The cross-scale dynamics in complex and adaptive ecosystems have been explained through a nested set of adaptive cycles operating at discrete scales in panarchy theory (Gunderson and Holling, 2001). In ecology, Schneider (2001) argues based on hierarchy theory that patterns in complex systems measured at small scales do not necessarily hold at larger scale, because processes, structure and variables seem to operate at discrete ranges of scale. Garmestani et al. (2009) have been able to detect scale-specific patterns with a discontinuity analysis.

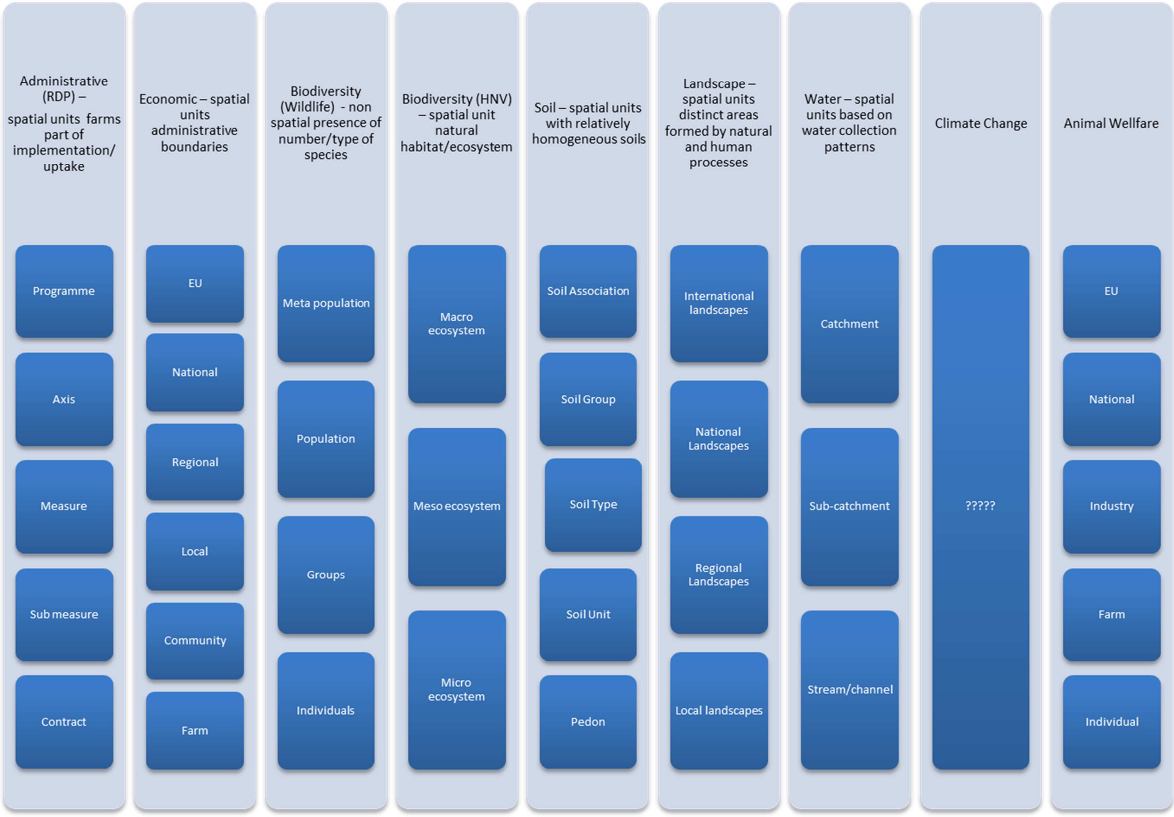

Figure 3 Scales in RDP modified after Cash (2006)

The advances in relation to understanding complex systems in this way are based on the dynamic processes in those ecosystems. In order to assess the impact of RDP measures on the environment, it is important that both the action and the measurement are placed in the appropriate context of the complex system. Hierarchy theory may suggest that the measurement/indicator should be at a higher level along the scale than the RDP action.

The definition of landscape by the European Landscape Convention is "*An area as perceived by people, whose character is the result of the action and interaction or natural and/or human factors*." Prager et al. (2012) argue that based on this definition 'landscape scale' refers to a spatial scale above the field-, farm- and local scale; and can be a catchment, an area of coherent landscape character or a sub-unit of a natural region. The importance of landscape for the delivery of local and regional ecosystem services is illustrated by Goldman et al. (2007) through an example of local and regional services, pollination and hydrologic services respectively (see Figure 4).

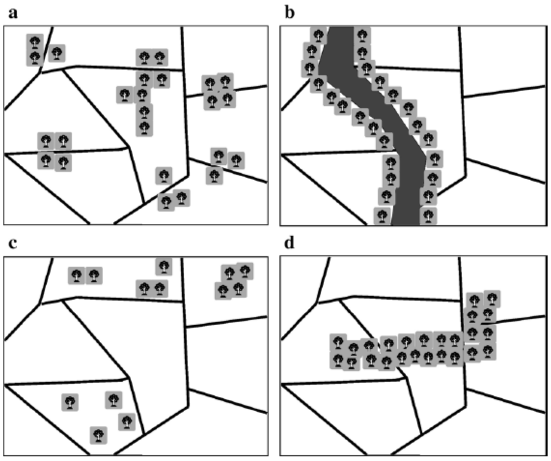

Figure 4 Four possible landscape configurations that could promote certain services: a) pollination, b) water purification, c) critical mass of participating land owners, d) landscape configuration (Source: Goldman et al., 2007).

3 Current Methods in Macro-level RDP Evaluation

The review of recent RDP evaluations in Austria, Finland, France, Germany, Great Britain (England and Scotland), Greece, Hungary, Italy, Lithuania and the Netherlands included the reported use of macro level (i.e. programme-level) assessment, which is reported here.

Priority was given to review reports in the following order:

1. 2007-2013 On-going evaluation reports and available ex-post evaluation results, cases with macro level reviewed
2. 2007-2013 Mid-term evaluation, cases with macro level reviewed
3. 2007-2013 Strategic Environmental Assessment, cases with macro level reviewed
4. 2000-2006 Ex-post evaluation reports, x cases with macro level reviewed
5. 2000-2006 Mid-term evaluation reports, x cases with a macro level reviewed
6. Other relevant RDP evaluation reports, x cases with a counterfactual reviewed.

We have collated the reported macro-level (programme) assesments explicitly mentioned in the evaluation reports. Thus the accuracy and specificity of the original evaluation reporting drives our review results. In Table 1 we have listed the number of evaluations with reported macro-level (RDP programme) assessment by affected public good and country.

More detailed results regarding the types of macro-level assessment methods used in evaluations are reported by public good in Tables 2 to 8. The report numbers refer to the numbers in brackets behind the evaluation reports in the list above. For identified cases we report on the type of environmental impact indicator, type of analysis, scale, causality, net impact and link between micro and macro. Where possible the type of method used for analysis is classified broadly into one of the **five categories**: multi-scale sampling (M), integrated/landscape (L), quantitative and statistical analyses (A), scaling (S) and qualitative (Q) approaches.

Table 1 Number of impact evaluations at a macro level by public good categories

Country	Climate	Water quality	Biodiversity(wildlife)	Biodiversity (HNV)	Soil	Landscape	Animal welfare
AT	-	2	1	-	-	2	-
DE	-	6	9	2	1	-	-
FI	-	-	-	-	-	-	-
FR	1	2	6	2	-	-	2
EL		-	5	9	-	-	-
HU		-	1	-	-	-	-
IT	3	8	-	1	-	6	1
LT	-	-	-	-	-	-	-
NL	-	-	-	-	-	-	-
UK	8	6	8	10	4	10	3
TOTAL	12	24	30	24	5	18	6

Tables 2 to 8 illustrate that the types of analysis used at a macro level are either qualitative or quantitative. There are very few examples of multiscale or integrated methodologies. In many cases there is no reported information regarding the scale, causality, net impact or the link between micro and macro assessment. The information is patchy and shows great variation between the different geographic areas.

Table 2 Analysis of climate-related impact assessment at the macro level

Country	Report	Measure or Programme	Impact indicator	Type of analysis	Temporal or Spatial	Causality	Net-impact assessed	Link micro macro
France	2	221	Stand IPCC for estimating N2O emmissions	A				
Italy (Veneto)	2	214, 221	Reduction of Nitrogen and CO2 emmissions	A				
Italy (Veneto)	4	H/I	Net carbon storage	A	Y		Y	
UK (Scotland)	2	111, 122, 214, 221, 223, 225, 227	Impact assessment is based on evaluation questions (by measure)	Q			Y	
UK (Scotland)	2	Programme	Impact assessment is based on evaluation questions (horizontal)	Q				
UK (England)		214, 221	Surplus nitrate/ha	Q	Y		N	

Table 3 Analysis of water quality related impact assessment at the macro level

Country	Report	Measure or Programme	Impact indicator	Type of analysis	Temporal or Spatial	Causality	Net-impact assessed	Link micro macro
Austria	2	214	Changes in gross nutrient balance	A	Y(t)		N	
Austria	2	Programme level	Improvement of water quality (GNB)	A				
DE (Baden Wuttenberg)	2	214 and others derived from 214 (NB1, NB3, ND2)	Stock density (LU)/ha)	Q	N	N	N	N
DE (Baden Wuttenberg)	2	Programme level	Water quality discussed with indicators such as N-depositions and N-surplus	Q			N	
DE (Lower Saxony)	2	214		A	Y(s)		N	Y
France	2	214	Excess in nitrate use	M	Y		y	

Country	Report	Measure or Programme	Impact indicator	Type of analysis	Temporal or Spatial	Causality	Net-impact assessed	Link micro macro
France	4	M	Nitrate reduction in surface water	A				
Italy (Puglia)	2	IRENA	Reduction of nitrate in ground water	A				
Italy (Puglia)	6	211, 212, 214, 221, 223, 226	Index LIM (Level of pollution Macrodescriptors)	A				
Italy (Veneto)	4	F	VI.I.B-1.2, VI.1.B-1.3, VI.1.C-1.1	A	Y(t)			N
UK (Scotland)	2	111, 214,216, 225, 227	Impact assessment is based on evaluation questions (by measure)	Q			Y	
UK (Scotland)	2	Programme level	Impact assessment is based on evaluation questions-horizontal	Q				
UK (England)	2	214	Surplus nutrient per hectare					Y

Table 4 Analysis of wildlife related impact assessment at the macro level

Country	Report	Measure or Programme	Impact indicator	Type of analysis	Temporal or Spatial	Causality	Net-impact assessed	Link micro macro
Austria	2	214	Farmland Bird Index	L	Y	Y	N	
DE (Baden-Wurtenberg)	4	212	Impact assessment is based on evaluation questions (by measure)	Q				
DE (Baden-Wurtenberg)	2	214	Crop diversity	A	N	N	N	
DE (Baden-Wurtenberg)	2	224	Impact assessment based on evaluation questions	Q			N	
DE (Baden-Wurtenberg)	4	212	Qualitative assessment / expert judgement based on comparison programme design and habitat requirements of bird species	Q	N		N	
DE (Baden-Wurtenberg)	4	214	Species diversity	A			N	
DE (Baden-Wurtenberg)	4	214	Number of species supported	A			N	
DE (Lower Saxony)	1	Programme	Farm land index	Q	N	Y	N	N
DE (Lower Saxony)	1	121	Change in grassland area	A	Y (t)		N	
DE (Lower Saxony)	2	214 (NAU/BAU B2 - result-oriented grassland extensification)	Number of indicator species	Q			N	
France	2	214	Recorded bird populations	A	Y			
France	4	F	Quantity & quality of bird species	A				
France	6	211, 212, 214, 216	Recorded bird populations	A	Y			
Greece	2	213, 214,221, 224, 226	Changes in extend of areas under successful management for improving biodiversity	A				
Hungary	4	214	Bird species abundance	A	Y(s)		Y	
UK (Scotland)	2	125, 212, 214, 221, 223, 225, 227	Impact assessment is based on evaluation questions (by measure)	Q			Y	
UK (Scotland)	2	Programme	Impact assessment is based on evaluation questions (horizontal)	Q				
UK (England)	2	221		Q				

Table 5Analysis of HNV related impact assessment at the macro level

Country	Report	Measure or Programme	Impact indicator	Type of analysis	Temporal or Spatial	Causality	Net-impact assessed	Link micro macro
DE (Baden Wurtenberg)	4	214	Impact assessment based on discussion	Q		Y	N	
DE (Lower Saxony)	1	Programme	HNV farmland area. This Indicator differentiates between different HNV areas and elements classified into different HNV types	L	Y	Y	N	N
France	2	214	HNV farmland areas	L				
France	4	F	HNV farmland areas – proportion of eligible farms receiving payment	L				
Greece	2	211, 212, 213, 214, 216, 221, 224, 226, 227	Maintenance of HNV farming areas	A				
Italy (Puglia)	4	F	HNV farm habitats	A				
Italy (Veneto)	4	F	HNV farm habitats	A				
UK (Scotland)	2	111, 125, 212,214, 216, 221,223, 225, 323	Impact assessment is based on evaluation questions (by measure)	Q				
UK (Scotland)	2	Programme	Impact assessment is based on evaluation questions (horizontal)	Q				
UK (England)	2	221		Q				

Table 6 Analysis of soil related impact assessment at the macro level

Country	Report	Measure or Programme	Impact indicator	Type of analysis	Temporal or Spatial	Causality	Net-impact assessed	Link micro macro
DE (Lower Saxony)	2	214 (C- organic farming)	Soil ersoion - estimation of the C factor for soil erosion	A	N	N	N	
UK (Scotland)	2	111, 214, 225, 227	Impact assessment is based on evaluation questions (by measure)	Q				

Table 7 Analysis of landscape related impact assessment at the macro level

Country	Report	Measure or Programme	Impact indicator	Type of analysis	Temporal or Spatial	Causality	Net-impact assessed	Link micro macro
Austria	2	211, 212	Stocking density per forage area	A		N		
Italy (Puglia)	2	214	Willingness to pay for conservation and landscape attributes	A				
Italy (Puglia)	4	H	VIII.2.B-2.2. Employment in the short/medium term outside holdings (logging, initial processing and marketing, and further local, small scale processing and marketing) directly or indirectly depending on assisted actions	A			N	
Italy (Puglia)	4	F	VI.3-2.1. Farmland under agreement contributing to perceptive/cognitive, in particular visual, differentiation (homogeneity/diversity) in the landscape (number of sites and hectares/ kilometres)	A			N	
Italy (Veneto)	2	214	'Willingness to pay' for the conservation of components and landscape attributes	A				
Italy (Veneto)	4	F	VI.3-2.1. Farmland under agreement contributing to perceptive/cognitive, in particular visual, differentiation (homogeneity/diversity) in the landscape (number of sites and hectares/ kilometres)	A			N	

Country	Report	Measure or Programme	Impact indicator	Type of analysis	Temporal or Spatial	Causality	Net-impact assessed	Link micro macro
Italy (Veneto)	4	H	VIII.2.B-3.1. Additional attractive/valuable area or sites due to assistance	A				
UK (Scotland)	2	111, 125, 212,214, 216, 221,223, 225, 323	Impact assessment is based on evaluation questions (by measure)	Q				
UK (Scotland)	2	Programme	Impact assessment is based on evaluation questions (horizontal)	Q				

Table 8 Analysis of animal welfare related impact assessment at the macro level

Country	Report	Measure or Programme	Impact indicator	Type of analysis	Temporal or Spatial	Causality	Net-impact assessed	Link micro macro
France	5	G	Proportion of friendly farming systems	S				
France	6	133	Individual aid for quality	S				
Italy (Veneto)	4	M	I.4-2.1. Share of assisted products sold with quality label (%); (a) of which EU-level labelling schemes (%)	S				
UK (Scotland)	2	212, 215	Impact assessment is based on evaluation questions (by measure)	Q				
UK (Scotland)	2	Programme	Impact assessment is based on evaluation questions (horizontal)	Q				

In general the results of the review of macro-level methodologies in the evaluation reports are patchy. In many cases the macro-level evaluation is based on qualitative assessment (Q) often using the evaluation questions (CMEF) or expert/evaluator assessment, which suggests an absence of evidence for causality between RDP and delivery of public goods. The reported quantitative or analytical methods (A) are almost as common as the qualitative methods. These methods seem to be simple tallies in many cases and do not explicitly and clearly include a disentanglement of single measures and overall programme effects from other factors. However there are a number examples of more complex and integrated methodologies (L) and multiscale approaches (M) that do aim to address this issue (Lower Saxony – HNV and water quality, Austria - wildlife, Veneto – landscape, and France – water quality).

The methodologies vary per public good. The reported methods include:

- Climate change – estimation and aggregation
- Water quality – impact coefficient, index for level of pollution of macrodescriptors (LIM), OECD methods, monitoring networks, and case studies
- Biodiversity wildlife – paired comparisons between areas participating and non-participaring

- Soil – modelling erosion loss
- Landscape – Shannon index and choice experiments
- Animal – disaggregate analysis

The method of paired comparisons between areas participating and non-participating in a particular RDP scheme is in some countries integrated in the monitoring programmes of biodiversity, providing relevant information for counterfactual analysis. However this type of evidence is not available in all EU nations and, in particular at macro level, requires approaches toseparate out the effect from other external factors on the measurements. Hence there is a demand for more integrated methodologies as reported by Austria and Lower Saxony which also enable the assessment of net environmental impacts at macro level.

Overall, the review of the recent evaluation reports confirmed the key challenges for macro level evaluations highlighted in the introduction and emphassied the need for the integration of new methodological developments.

4 Inventory of Relevant Methodologies

In a recent review of evaluation methods for RDP, in an attempt to find an alternative approach to RDP evaluation, Terluin and Roza (2010) included 22 different evaluation methods for individual rural development measures. They classified the evaluation methodologies into five different approaches: CMEF-type approach, tally approach, econometric approach, modelling approach and mixed case-study approach. While the review considered the benefits and limitations of these methodological approaches, they are single measure, largely farm-based and only to a very limited extent include the complexity of scale and the delivery of public goods.

The emphasis of this current review is specifically on methods for a macro-level assessment of the environmental impact of the RDP and its key challenges: determination of true causation, aggregation of all economic, social and environmental net-impacts of RDP, and a cost-benefit analysis (Michalek, 2012). Therefore, a modified classification of approaches has been chosen for the methodologies to reflect the demand for methods that can address complex challenges: statistical approach, hierarchical and multi-scale approach, spatial analytical approach, multi-criteria approach, and integrated/landscape approach.

4.1 Statistical Approaches

Statistical approaches have been used to assess the relationship between farm based activities or decision-making to climate change and biodiversity. Reidsma and Ewert (2008) have used a vulnerability analysis based on the Pearson Correlation, the Shannon-Weaver Index and a linear regression model to assess the impact of farm diversity on the vulnerability to climate change. A multi-level analysis of yield and income responses to climate variability based on a General Linear Model has been used for estimating and analysis of trends. With this methodology Reidsma et al. (2009) set out to explain the relationship between climate and management impacts on farmers' income and crop yield at two different scale levels (farm and region) using FADN and climate change data. These types of analysis can potentially provide evidence of causality. However, in relation to the complex nature of the causality between RDP measures and public goods, these methods have some relevance but these may be limited in the environmental impact evaluation. Another type of statistical approach is deployed in relation to scaling. Scaling is important for the integration of data that is needed to assess the impact of agri-environmental measures on public goods. This may involve either up- or down-scaling, depending of the difference between scale of indicator(s) and scale of analysis. Scaling methods can be classified by manipulation of data and manipulation of models (Ewert et al., 2011).

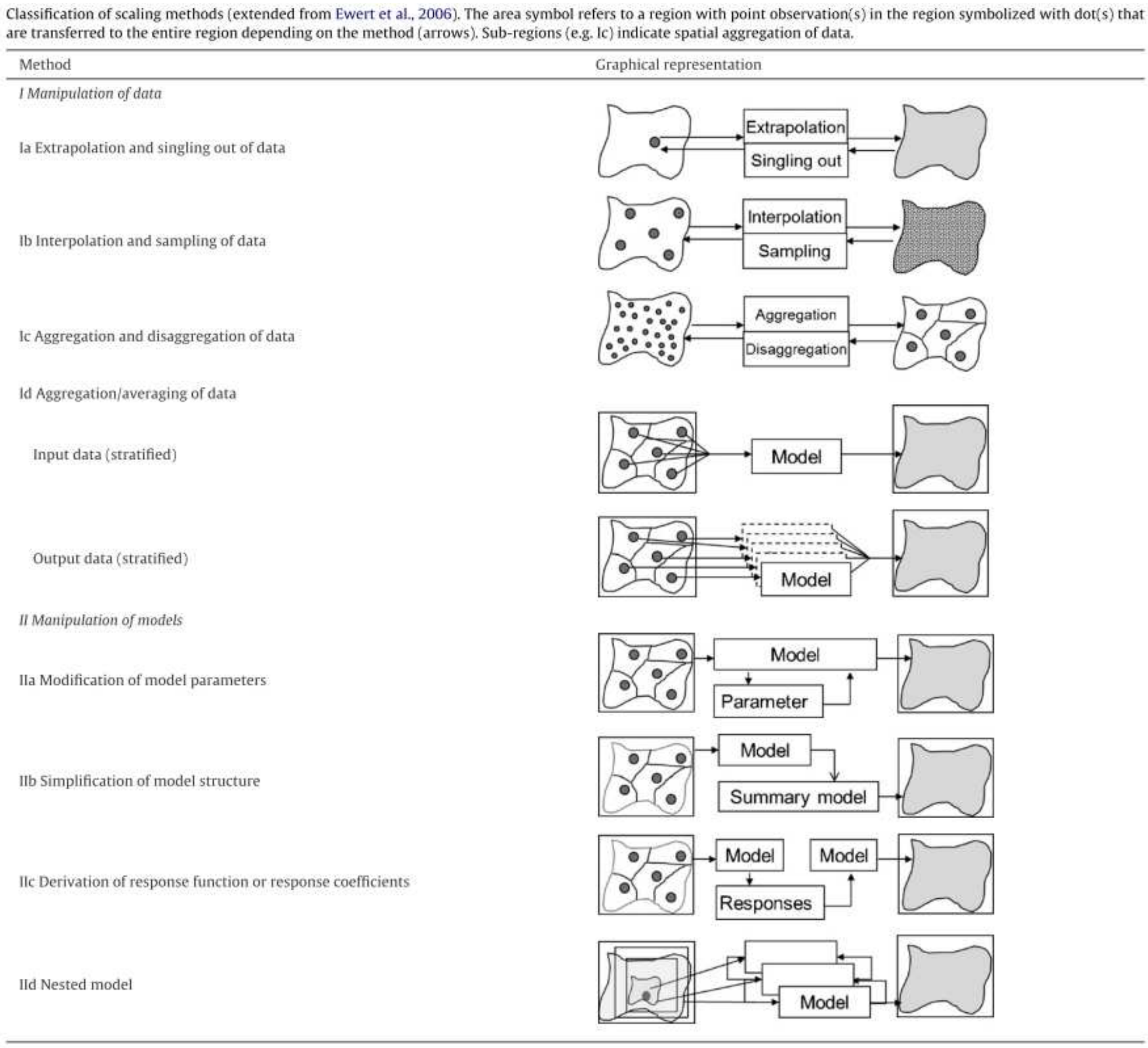

Figure 5 Classification of scaling methods from Ewert et al. (2011)

Araujo and Thuiller (2005) present a stepwise downscaling process (generalised additive modelling) using interpolated European data (10x10km) which were first aggregated to a coarse level (50x50km) and then modelled back to their original resolution. With this process of up and downscaling they illustrated that, despite uncertainties, downscaling may prove useful for the identification of reserves more meaningfully related to local patterns of environmental variation. However this method is limited to regions that are data-rich. Righi et al. (2011) and Andersen et al. (2007) use the development of a farm typology framework as a means to up-scale farm-level impact assessment to regional level. For each farm type they used a simulation to model an average farm, and field data from representative farms to assess the impact of farm-level strategies. This systematic approach of characterising farms in typologies is interesting and allows for an alternative to simple aggregation of variables at farm level. However the reported research and the impact assessment did not include any public good. Based on a homogeneous agro-ecological regions, Small Agricultural Region, Teillard et al (2012) used a two-step multinomial regression to estimate the input cost per hectare value and they used local Moran for the aggregation of farm intensity.

The identification of relatively homogeneous agricultural areas is thought to have unique responses and possibly more effective in relation to delivering specific RDP measures, particularly because, while the aggregated clusters are not consistent with administrative boundaries, they may link more closely to the delivery of public good by agricultural areas.

Given the multi-scale and multi-level nature of the macro-level assessment of the impact of RDP and the recognised mismatch between economic and ecological scales, the process of scaling and the use of farming typologies are likely to be important components of the methodological framework for ENVIEVAL. They provide potentially a mechanism to link the micro- and macro-level assessments.

4.2 Multi-scale/Hierarchical Approaches

Baker et al. (2011) assessed the impact of agricultural management on bird populations at three different spatial resolutions. They used the Breeding Bird Survey (BBS) data and Environmental Stewardship to analyse the 1km^2 cell of the BBS as well as two buffer areas around the cell of 9km^2 (3 x 3) and 25 km^2 (5 x 5). With a loglinear approach they modelled the change in expected abundance between consecutive years which incorporated the effects of spatio-temporal covariates. They suggest that for the reversal of farmland bird numbers there is a need for management response at a different temporal and spatial scale from what they call broad-and-shallow AES approach.

An alternative method to assess the influence of human practices on biodiversity patterns by Deconchat et al. (2007) used an interdisciplinary nested sampling approach to analyse the interaction of social and ecological processes around forested areas in rural landscapes. Although these authors acknowledge the difference in sampling methodology and scale between relevant scientific disciplines, the integration of the sampled data was facilitated through selection of common sites, synchronisation of measurements, establishing collective protocols and sharing data management tools. They argue that "*Splitting the project into sub-parts is not a reductive approach if different scales are studied simultaneously and if the methodologies are chosen to be compatible*". However this does require a high level of inherent coherence and coordination. This particular methodology is highly time and resource demanding and hence not realistic for impact assessments of RDP. However, a nested approach can prove relevant in dealing with cross-scale and cross-level assessments, which are necessary for the RDP.

A hierarchical sampling design with systematic and consistent sampling method across levels supports a multiscale assessment of impact of farming on biodiversity (Gabriel et al., 2006). Principal Component Analysis on 30 variables describing climate, topography, land use, socio-economic and soil conditions was used to select matching landscapes, 10 x 10 km landscapes with either high or low organic farming. Landscapes were paired on the basis of similar environmental conditions but contrasting in surrounding presence of organic farming (Gabriel et al., 2010). Within each landscape comparable organic and conventional farms were selected, and within each farm three cereal fields and three grass fields were selected (Gabriel et al., 2010). This method is a very systematic approach to handling scale which is necessary to fully understand the impact of RDP on a particular biodiversity. The sampling design is robust but is expected to be time and resource intensive which may pose a constraint on the EU-wide implementation. However, due to the landscape characterisation, the actual sampling may be more effective.

A multi-scale assessment modelling framework (SEAMLESS) assesses the impact of agricultural policy measures on the farm business and incorporates a process of scaling that integrates models of micro and macro level. The framework uses the concept of actor-action-environment-condition and related model linkages to connect farm, regional and EU levels. The creation of homogenous spatial units and typologies for describing farms and the environmental context facilitate a systematic aggregation of detail and to address the scale mismatch by integrating the socio-economic and biophysical data (Report_14) Scaling methods are essential components to link between different scale levels and models (Ewert et al., 2011, see Figure 5).

While this methodology incorporates environmental components, it does not assess the impact of the agricultural management specifically on public goods. The method of aggregation and disaggregation however is potentially relevant for the methodological framework to be proposed by ENVIEVAL.

The multiple scale and hierarchical methods have an important role in the environmental impact assessment of RDP as they are able to provide evidence of causality and a systematic approach to linking micro and macro level assessment.

4.3 Spatial Analytical Approach

Spatial econometric modelling approaches are receiving increasing attention for their use in assessing impacts of CAP policies (e.g. Reinhard et al., 2013; Montresor et al., 2010; Pecci & Sassi, 2008). The most recent development and application of spatial econometric models for the evaluation of environmental (and socio-economic) impacts of RDPs was carried out in the SPARD project (Reinhard et al., 2013; Reinhard & Linderhof, 2013).

Traditional or non-spatial econometrics largely ignores the spatial dependence and spatial heterogeneity of data sets used in econometric models. Spatial econometrics incorporates those issues in the estimation process based on spatial data. Spatial econometrics is characterised by: spatial dependence between sample data observations at various points in the Cartesian plane, and spatial heterogeneity that arises from relationships or model parameters that vary with the sample data as one moves over the Cartesian plane. Given that many rural development measures and their environmental impacts have an explicit spatial dimension further emphasises the need to use spatial econometric techniques for the quantitative analysis of environmental impacts of rural development measures. Reinhard et al. (2013) summarise that, in more advanced spatial econometric models, two main approaches are applied to consider situations:

a) where the outcome in one region is affected by the outcome in neighbouring regions (a spatial lag model);
b) where the outcome in one region is affected by unknown characteristics of the neighbouring regions (a spatial error model).

Reinhard et al. (2013) developed and tested spatial lag models (using a Gabriel weight matrix) for the assessment of impacts of agri-environmental measures on water quality and biodiversity. N-surplus was used as an indicator for water quality, and a HNV index based on Paracchini and Britz (2010) was used as a proxy indicator for biodiversity. The impact analysis was tested at NUTS0 level for water quality and NUTS1 level for HNV farming. The data and estimation results show that spatial correlation is present but the spatial econometric models applied are not preferred over the a-spatial models. The authors largely explain this with the lack of disaggregated data on AEM spending in relation to specific environmental objectives and on the environmental impact indicators, which hindered model testing at more suitable levels such as NUTS2 or NUTS3. This highlights that data requirements are stringent for spatial econometric analysis.

Generally, Reinhard and Linderhof (2013) conclude that the tested spatial econometric models provide a suitable method to assess environmental impacts of RDPs at macro level which allows the incorporation of counterfactuals through analysing regions with different spending on the measures and different development trajectories of biodiversity and water quality. Micro-level impact analysis was not feasible due to lack of data on environmental impact indicators. A consolidated data base on impact and baseline indicators as well as other general trends is a prerequisite. Overall, the results of the SPARD project suggest that further exploration of the data requirements and the feasibility of testing spatial econometric models in the ENVIEVAL case studies for specific public goods should be pursued. In addition, the suitability of spatial econometric models to assess programme impacts could be explored (SPARD 'only' tested the models for one measure) and potentially be tested in the case studies. The case studies would test the suitability of spatial econometric models to address the challenge to assess net environmental impacts of rural development measures and programmes at macro level.

Gimona and van der Horst (2009) have shown the potential of landscape indices for environmental monitoring using CORINE Land Cover data. They selected nine indices which were calculated using both local detailed data and CLC2000 for 20 10x10 km tiles. They identified the first five principal components which they used as independent variables in linear regression models. They subsequently selected the best model using Multivariate Adaptive Regression-based Splines (MARS). This method has not yet been used beyond the case study area and it has not proven a strong link between the landscape indices and biodiversity. However, this method has illustrated the value of landscape indices for monitoring change in landscape that can assess change in the context of RDP and the benefit of targeted implementation of RDP measures. Concepcion et al. (2008) conducted a similar study but instead of a transect they focused specifically on the effectiveness of AES in the context of landscape complexity in three agricultural areas. The research included fieldwork in seven pairs of cereal fields in each of the three regions (measurement of species richness) and aerial photography interpretation (measurement of landscape metrics). The results from this study also conclude that the effectiveness of RDP measures are dependent on the complexity of the surrounding areas, which emphasises the need and value of landscape metrics and spatial analytical analysis for assessing the impact of RDP measures.

The above examples show the potential of landscape metrics in the impact assessment of RDP on biodiversity. However an example of measuring multiple public goods more generally is illustrated by the evaluation of an afforestation scheme developed by mapping multifunctional hotspots based on benefit maps and potential for public goods (biodiversity, visual amenity and woodland recreation) (Gimona and van der Horst, 2007). Based on benefit maps (i.e. potential biodiversity map, potential visual amenity map and potential on-site recreation map) multifunctional hotspots were identified as areas which had consistently high scores in different weighting schemes which were used to combine the benefit maps. A comparison between the multifunctional hotspots and Farm Woodland Premium Scheme (FWPS) illustrated the potential for this approach to assess the impact of the FWPS in meeting the policy objectives of public goods (biodiversity, amenity and recreation).

The impacts on public goods of potential spatial arrangements at the landscape scale can be explored using spatio-temporal modelling of land cover, land use or cropping systems. The landscape simulations can be based on stochastic processes and rule-based constraints allowing the integration of case study particularities derived from field scale (e.g. IACS to characterise cropping systems) within the context of regulatory policy or strategic plans. The simulations can be set up at one or multiple scales as relevant (e.g. landscape character unit, water catchment, NUTS level; Castellazzi et al., 2010).

Specifically for a macro-level assessment, spatial analytical methods are valuable because they accept the presence of spatial correlation in the indicators, have developed means of measuring indicators in their spatial context which is particularly important in relation to public goods.

4.4 Multi-Criteria Approaches

Interviews among participating and non-participating farmers have been used by Primdahl et al. (2003) to assess the effects of agri-environmental schemes based on 12 indicators. They found that, although the 12 indicators worked, follow-up among the sampled farms is necessary to secure an effective long-term monitoring plan. The key limitation of this method is that it has no spatial and/or scalar component, nor are there hard measurements of these improvements.

In general multi-criteria approaches provide a systematic approach for such a comparison. It includes a wide range of different approaches which include Multi Criteria Analysis (MCA), Life Cycle Analysis (LCA), Ecological Footprint (EF), Agri-Environmental Footprint (AEF) and Cost Benefit Analysis (CBA). These methods have in common that they compare different alternative options against a range of criteria; through a weighted sum methodology, a classification of alternatives is obtained and the final results can be ranked. MCA can be used to compare large sets of relevant criteria which can then be adjusted during assessment (e.g. Koo and O' Connell, 2006; Dujmovic and De Tre, 2011; Mattsson et al., 2000). The method can be used in a spatial context by applying an effect or decision-making matrix, set of criterion weights and decision rules to individual grid cells in the study area. The decision rule can generate site-specific land-use optimisations. Boggia and Cortina (2010) used the method to aggregate indicator data into two indexes for territorial level: an environmental index and a socioeconomic index. They argue that, rather than further aggregating the two indexes, an interpretation based on crossing values allows each dimension to retain its independence.

Life Cycle Analysis enables consideration of the life cycle consequences of a combination of effects of RDPs, and their individual components, with respect to public goods. The translation of life cycle information into area demand is termed 'footprinting', examples including Ecological Footprints (Rees, 1992), and Carbon Footprints. Thus, in the context of this review, footprinting methodologies are most appropriate, rather than 'true' LCA, as they allow for indicators at multiple scales to be addressed. All material and energy consumption of an economic unit, e.g. household, community, region, nation, is compiled and converted into the land and water areas that are necessary to supply this unit with the material and energy as well as to cope with its waste (Stoeglehner and Narodoslawsky, 2008). Weaknesses in the approach that have been highlighted include the fact that the original footprinting model had no policy advisory component and it was also not possible to distinguish between different qualities of material and energy flows. Furthermore, footprinting is said to have an anti-trade bias (van den Bergh and Verbruggen, 1999). However, individual authors have made specific modifications to the footprinting model that in many respects overcome these drawbacks (Ferguson, 1999).

These approaches seem to be most commonly used for ex-ante assessment rather than ex-post. However, in order for that to work, the defined criteria will need to include clear impact-related criteria that incorporate evidence of causality against which the measures are assessed. This means that, although the approach may be successful in handling multiple criteria, its use for an ex-post assessment will rely on the quality of the definition of the criteria.

The work on agri-environmental footprinting has demonstrated the value of multi-criteria analysis for the evaluation of environmental impacts on individual farms (Purvis et al., 2009). This method has shown to be able to assess change in the Agri-environmental Footprint Index (AFI) over time and to compare participating and non-participating farms. These types of results can be aggregated to higher levels on the scale (RDP programme/macro level). However, in the context of the landscape scale, it is important for the criteria to include landscape metrics as illustrated in section 4.3 in order to identify the impact of RDP in relation to the wider context. Research in relation to the biodiversity (HNV) has shown that the spatial context of implemented environmental measures is important for its wider impact.

In general, multi-criteria analysis provides a systematic and holistic approach to the comparison of the impact of individual as well as collective measures. It means that the methodology, in particular ecological footprinting, is able to address one of the key challenges of the RDP impact assessment, i.e. assessing the multiple measures and indicators. However, extending the AFI method with spatial analytical approaches will need to be considered for use in the methodological framework for ENVIEVAL.

4.5 Integrated/Landscape Approach

Increasing recognition of the complexity of coupled socio-ecological systems has motivated the development of integrated research frameworks that address the interactions between causal factors and outcomes at micro and macro levels, i.e. field, farm, regional, national and international levels (van Ittersum and Brouwer, 2010). Although the resulting models have many potential advantages, the complexity of 'integrated assessment' poses several outstanding conceptual and methodological issues. Moreover, their success has not been assessed by any in-depth economic analysis. An integrated assessment framework such as SEAMLESS (van Ittersum et al., 2008) includes different aspects (Ewert et al., 2009) by linking or coupling multiple models for macroeconomic, microeconomic (or behavioural) and biophysical aspects of the socio-environmental systems. While this integrated assessment framework is a modelling approach focussed specifically on agricultural land management and RDP, there are other integrated approaches which may be of relevance for the impact assessment in the context of public goods.

For example, Ferrier and Drielsma (2010) propose a flexible whole-landscape modelling framework. This framework for integrating multiple pattern and process-related factors into biodiversity conservation assessment is used by Williams et al. (2012) to develop and assess the effectiveness of conservation incentive payments. A whole-landscape metric for the assessment of biodiversity conservation has been developed based on a biodiversity assessment through habitat connectivity and ecosystem complementarity as well as a tender evaluation of conservation instruments.

In addition to these quantitative approaches, there are qualitative integrated approaches to assess the impact of RDP on public goods. Roth and Schwabe (2003) describe a participatory approach based on the role of agricultural land in the wider landscape through plans agreed with relevant stakeholders that secure a future for both agriculture and public goods. Target values for ecologically, culturally and regionally valuable areas are calculated per natural region/landscape character on the basis of soil, water, presence of valued habitats and habitat networks as well as landscape analysis. The pros are that this assesses all the public goods and the progress that has been made. It sets a target as well as a clear measure. The cons are that it is very intensive to set up due to data demands, as well as the stakeholder process and it may not be so easily scaled up to more abstract levels; a generalisation has been made but that is still only to regional level rather than (inter)national level.

Carey et al. (2003) and Carey et al. (2005) use a multidisciplinary expert panel to conduct a multidimensional assessment of two RDP measures (Countryside Stewardship Scheme, CSS, and Environmentally Sensitive Areas, ESA). The panel was to assess two agri-environmental schemes based on the content of the agreements/contracts and the farm's landscape context. The panel judged the individual agreements on agreement negotiation, appropriateness, environmental effectiveness, compliance and side effects as well as additionality (a kind of counterfactual assessment). Their results highlight the difference between the design and functioning of the two schemes and their ability to identify the scope for application of inappropriate management prescriptions. In this type of context conjoint analysis (CA) can be used to identify attributes of environmental management/policy options (e.g. land management options) and evaluate trade-offs between options, using different means for representing the data (e.g. landscape visualisations between counterfactual and scenarios of change), such as landscape impacts, GHG emissions (Sheppard, 2005).

Midmore et al. (2008) and Terluin and Berkhout (2011) used and presented a mixed case study approach for the evaluation of rural development programmes. The mixed case study approach aims at exploring patterns, which provide support for explanations of causal relationships and seeks to explain how rural development policies interact with other policy impacts and the governance framework which delivers support. It combines a quantitative analysis of input and output indicators at measure (micro) level with qualitative methods at programme (macro) level. The mixed case study approach fosters a better understanding of the causal-relationships and can provide a cost-effective evaluation method in situations with limited data availability for quantitative impact assessments.

5 Discussion

The results from the review show clearly that a lot of methodological progress has been made in recent years to improve a range of aspects regarding the environmental impact assessment of RDP measures, although many methods to date are only applied to ex-ante rather than ex-post assessment. The most important methodological developments are the advances made in relation to multi-criteria, spatial analytical approaches and integrated approaches, as well as efforts made to address the scale mismatch between economic and ecological/natural sciences. These developments are able to contribute to addressing the challenges posed by the demand for measuring the impact of RDP activities/investment on the delivery of public goods.

There is evidence that just tallying the implementation of the measures does not equate to successful delivery of public goods and that it does matter to the success of RDP where and how measures are implemented. At the same time the issue of scale in relation to measuring impact has also been shown to be important. Evidence of cause and effect between RDP and public goods is largely limited to biodiversity (wildlife) and water quality.

The methodologies used for the macro (programme level) assessment is patchy and in many cases is not actually making an assessment at programme level but at measure level. Multi-criteria analysis, in particular agri-environmental footprinting (Purvis et al., 2009), is the most promising approach to assess the complexity of assessing multiple measures. However a key challenge will be to take this meaningfully from farm-level to regional and national level. In this context, lessons from other methodological approaches for scaling (Ewert et al., 2011) and conceptual integration may be valuable (van Ittersum and Brouwer, 2010; Ferrier and Dreilsma, 2010).

Based on the developments towards integrated assessments, macro-level assessment should focus not just on the overall programme assessment (i.e. effectiveness of investment), but more to the wider environmental impact which can be at different geographic scales. This does require the inclusion of spatially-explicit analysis at appropriate scale levels. Previous EU projects have addressed a number of relevant issues in the effort for integrated analysis.

Recent developments in spatial econometrics in principal address the need to include spatially-explicit analysis at appropriate scale levels. The case study applications of new spatial econometric models in the SPARD project have indicated that such models have the potential to assess and quantify net-impacts at macro level, which was confirmed as one of the key challenge for macro-level evaluations in the stakeholder interviews and workshop discussions. Including spatial econometric models in the ENVIEVAL case studies would enable the testing of the suitability of spatial econometric models to address the challenge to assess net environmental impacts of rural development measures and programmes at macro level. However, a potential limitation for the use spatial econometric models is their large and complex data requirements. The data requirements and availability need to be assessed in more detail in the next task to ensure that feasible case study areas with sufficient data infrastructure can be selected and validated with the SRG.

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

The main gaps in the current knowledge and candidate methods for the case study testing are synthesised below and Table 9 provides an overview of the different methods reviewed including key aspects such as addressed challenges in evaluations mechanisms to implement counterfactuals, mechanisms to link with micro level evaluations, and data requirements.

Main gap in the current knowledge:

- limited evidence of causality between the RDP objectives and indicators
- limited evidence of the ability of indicators to measure impact across and within scales and levels
- the need to incorporate the spatial context around participating areas.

Candidate methodologies:

- Increasing evidence base
    - Hierarchical approaches
        - systematic and consistent sampling method (Gabriel et al., 2006, 2010; Concepcion et al., 2012; Deconchant et al., 2007)
        - use of typologies (Righi et al., 2011; Andersen et al., 2007; Teillard et al., 2012)
        - scaling (up- and downscaling - Ewert et al., 2011)
    - Spatial analytical approaches –
        - landscape metrics - the flexible whole-landscape modelling framework by Ferrier and Drielsma (2010)
        - spatial econometrics – by Reinhard et al., 2013; Reinhard and Linderhof, 2013
- Assessment
    - Multi-criteria analysis – agri-ecological footprint
    - Spatial econometrics
    - Landscape zoning and multi-functional hotspots

Table 9 Overview of different methods reviewed

Method	Key methodological aspect	Addressed challenge(s) or improvement of evaluation	Mechanisms to implement counterfactuals	Mechanisms to link with micro level	Data requirements	Other key requirement	Tested / applied for which public good / environmental aspect	Source(s)
Statistical	Multi-scale/dimension trend analysis	determination of true causation		Scaling	FADN in HARM regions		Climate	Reidsma et al. (2009)
	GLM Modelling	determination of true causation		Scaling	Species abundance data (butterflies and farmland birds)		Biodiversity	Aviron et al. (2007); Davey et al. (2010); Gottschalk et al. (2007)
	Scaling	data/model integration for aggregation RDP impacts			MARS, FADN, Eurostat, GTAP databases			Ewert et al. (2011)
	Down scaling using generalised additive modelling (GAM)	data integration for aggregation RDP impacts		down-scaling	Species abundance, landcover and climate data		Biodiversity	Araujo and Thuiller (2005)
	Up scaling	data integration for aggregation RDP impacts		field data capture	farm management data, environmental data		Soil	Righi et al. (2011)
Hierarchical	Hierarchical sampling	assessment of Impact on the wider context (public goods) – true causation		field level observations/ data	Climate, topographic, land ue, social/economic and soil conditions as well as field data		Biodiversity	Gabriel (2006, 2010); Concepcion et al. (2012); Deconchat et al. (2007)
	Cluster analysis, regression	assessment of Impact on the wider context (public goods) – true causation		aggregation of micro level to production type	FADN,		Biodiversity and biodiversity (HNV)	Amano et al. (2011); Teillard et al. (2012)
	Statistical modelling at different spatial	assessment of Impact on the wider context (public goods) – true			BSS and ES data		Biodiversity	Baker et al. (2011)

Method	Key methodological aspect	Addressed challenge(s) or improvement of evaluation	Mechanisms to implement counterfactuals	Mechanisms to link with micro level	Data requirements	Other key requirement	Tested / applied for which public good / environmental aspect	Source(s)
	resolutions	causation						
Spatial analytical	Spatial econometrics	assessment of net impact on the wider context (public goods) – true causation	Comparison of regions with different spending and different developments in public goods	aggregation of farmlevel data	CMEF data, Cambridge Econometrics database		Biodiversity, water quality and tourism	Reinhard et al. (2013); Montresor, (2010); Pecci and Sassi (2008); Reinhard and Linderhof, (2013)
	Landscape analysis Landscape metrics and zoning	assessment of Impact on the wider context (public goods) – true causation	Spatial rather than temporal	Field level observations/ data	Public good monitoring data/maps, RS data and field/farm data (RDP implementation)		All public goods, but specifically biodiversity (HNV) and landscape	Concepcion et al. (2008); Gimona et al (2009); Gimona and van der Horst (2007); Van der Horst and Gimona (2005); Castellazzi et al. (2010)
Multi-criteria	Survey/panel Knowledge elicitation	aggregation of all economic, social & environmental effects of RDP		Farmer interviews	Survey or panel data		Water and soil quality	Primdahl et al. (2003)
	Multi-Criteria Analysis	aggregation of all economic, social & environmental effects of RDP			Depending on the criteria identified for the assessment		All public goods	Koo and O’Connell (2006); Boggia and Cortina (2010);
	Life Cycle Analysis	aggregation of all economic, social & environmental effects of RDP		Farmlevel assessment included in assessment	Depending on the criteria/indicators used.		All public goods	Rees (1992); Van den Bergh and Verbruggen (1999); Ferguson (1999); Stoeglehner and Narodoslawsky (2008)
	Ecological footprint analysis	aggregation of all economic, social & environmental effects of RDP		Farmlevel assessment included in assessment			landscape quality, biodiversity and natural resources	Purvis et al. (2009)
Integrate	Expert	aggregation of all	Ask	none	Expert or stakeholder		biodiversity,	Carey et al. (2005, 2010)

Method	Key methodological aspect	Addressed challenge(s) or improvement of evaluation	Mechanisms to implement counterfactuals	Mechanisms to link with micro level	Data requirements	Other key requirement	Tested / applied for which public good / environmental aspect	Source(s)
d	assessment/ Knowledge elicitation	economic, social & environmental effects of RDP			opinion		landscape and cultural services	
	Mixed case studies	aggregation of all economic, social & environmental effects of RDP			Quantitative (input and output indicators) and qualitative data.	Focus on impact policy objective	RPD environmental impact (no public goods)	Terluin and Berkhout (2011); Midmore et al. (2008)
	Model integration	aggregation of all economic, social & environmental effects of RDP		Scaling, typology	MARS, FADN, Eurostat, GTAP databases			Van Ittersum and Brouwer (2010); van Ittersum et al. (2008); Ewert et al. (2009); Ferrier and Drielsma (2010); Williams et al. (2012)
	Mapping landscape functions and services	Quantification of change in the value of a public good		Disaggregation to farm level	Land cover, policy documents, habitat data, agricultural production data, tourism data		All public goods	Willemen et al. (2008, 2010);
	Adaptive Management (?)	Quantification of change in the value of a public good per natural region/landscape character		none	Farm level data and aerial photographs		Landscape and biodiversity	Roth and Schwabe (2003)

Bibliography

Adger W.N., Brown K., Tomkins E.L. (2006) The political economy of cross-scale networks in resource co-management. *Ecology and Society* 19(2): 9.

Amano T., Kusumoto Y., Okamura H. et al (2011) A macro-scale perspective on within-farm management: how climate and topography alter the effect of farming practices. *Ecology Letters* 14(12): 1263-1272.

Araújo M., Thuiller W. (2005) Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography* 14: 17–30.

Aviron S., Jeanneret P., Schüpbach B., Herzog. F. (2007) Effects of agri-environmental measures, site and landscape conditions on butterfly diversity of Swiss grassland. *Agriculture, Ecosystems & Environment* 122: 295–304.

Baker D.J., Freeman S.N., Grice P. V., Siriwardena G.M. (2012) Landscape-scale responses of birds to agri-environment management: a test of the English Environmental Stewardship scheme. *Journal of Applied Ecology* 49, 871–882.

Bezlepkina I., Reidsma P., Sieber S., Helming K. (2011) Integrated assessment of sustainability of agricultural systems and land use: Methods, tools and applications. *Agricultural Systems* 104(2): 105-109.

Boggia A., Cortina C. (2010) Measuring sustainable development using a multi-criteria model: a case study. *Journal of Environmental Management* 91(11): 2301-2306.

Carey P.D., Manchester S.J., Firbank L.G. (2005) Performance of two agri-environment schemes in England□: a comparison of ecological and multi-disciplinary evaluations. Agricultural Ecosystems and Environment 108: 178–188.

Carey P.D., Short C., Morris C., Hunt J., Priscott A., Davis M., Finch C., Curry N., Little W., Winter M., Parkin A., Firbank L.G. (2003) The multi-disciplinary evaluation of a national agri-environment scheme. *Journal of Environmental Management* 69: 71–91.

Cash D.W., Adger W., Berkes F., Garden P., Lebel L., Olsson L., P., Pritchard L., Young O. (2006) Scale and cross-scale dynamics: governance and information in a multilevel world. *Ecology and Society* 11(2).

Castellazzi M., Matthews J., Angevin F., Sausse C., Wood G., Burgess P., Brown I., Conrad K., Perry J. (2010) Simulation scenarios of spatio-temporal arrangement of crops at the landscape scale, an application of the LandSFACTS tool. *Environmental Modelling and Software* 25: 1881-1889.

Concepción E.D., Díaz M., Kleijn D., Báldi A., Batáry P., Clough Y., Gabriel D., Herzog F., Holzschuh A., Knop E., Marshall E.J.P., Tscharntke T., Verhulst J. (2012) Interactive effects of landscape context constrain the effectiveness of local agri-environmental management. *Journal of Applied Ecology* 49, 695–705.

Concepción E. D., Díaz M. (2011) Field, landscape and regional effects of farmland management on specialist open-land birds: Does body size matter? *Agriculture, Ecosystems & Environment* 142(3–4): 303-310.

Concepción E., Díaz M., Baquero R. (2008) Effects of landscape complexity on the ecological effectiveness of agri-environment schemes. *Landscape Ecology* 23(2): 135-148.

Cooper T., Baldock D., Rayment M., Kuhmonen T., Terluin I., Swales V., Poux X., Zakeossian D., Farmer M. (2006) An Evaluation of the Less Favoured Area Measure in the 25 Member States of the European Union. A report prepared by the Institute for European Environmental Policy for DG Agriculture.

Davey C.M., Vickery J.A., Boatman N.D., Chamberlain D.E., Parry H.R., Siriwardena G.M. (2010) Assessing the impact of Entry Level Stewardship on lowland farmland birds in England. *Ibis* 152: 459–474.

Deconchat M., Gibon A., Cabanettes A., du Bus de Warnaffe G., Hewison M., Garine E., Gavaland A., Lacombe J-P., Ladet S., Monteil C., Ouin A., Sarthou J-P., Sourdril A., BalentG. (2007) How to set up a research framework to analyze social–ecological interactive processes in a rural landscape. *Ecology and Society* 12(1): 15. [online] URL: http://www.ecologyandsociety.org/vol12/iss1/art15/

Dujmovic J., De Tre G. (2011) Multicriteria methods and logic aggregation in suitability maps. *International Journal of Intelligent Systems* 26(10): 971-1001.

Dwyer J., Ward N., Lowe P., Baldock D. (2007) European rural development under the Common Agricultural Policy's 'Second Pillar': Institutional conservatism and innovation. *Regional Studies* 41(7): 873-887.

European Commission (2006) Handbook on Common Monitoring and Evaluation Framework 2006. http://ec.europa.eu/agriculture/rurdev/eval/

European Court of Auditors (2011) Is agri-environment support well-designed and managed? Special Report No. 7/2011.

Ewert F., Van Ittersum M.K., Heckelei T., Therond O., Bezlepkina I., Andersen E. (2011) Scale changes and model linking methods for integrated assessment of agri-environmental systems. *Agriculture, Ecosystems & Environment* 142: 6–17.

Ewert F., van Ittersum M.K., Bezlepkina I., Therond O., Andersen E., Belhouchette H., Bockstaller C., Brouwer F., Heckelei T., Janssen S., Knapen R., Kuiper M., Louhichi K., Olsson J.A., Turpin N., Wery J., Wien J.E., Wolf J. (2009) A methodology for integrated assessment of policy impacts in agriculture. *Environmental Science & Policy* 12: 546-561.

Fekete A., Damn M., Birkmann J. (2010) Scales as a challenge for vulnerability assessment. *Natural Hazards* 55: 729-747.

Ferguson A.R.B. (1999) The essence of ecological footprints. *Ecological Economics* 31: 318-319.

Ferrier S., Drielsma M. (2010) Synthesis of pattern and process in biodiversity conservation assessment: a flexible whole-landscape modelling framework. *Diversity and Distributions* 16(3): 386-402.

Gabriel D., Roschewitz I., Tscharntke T., Thies C. (2006) Beta diversity at different spatial scales: plant communities in organic and conventional agriculture. Ecological applications : a publication of the *Ecological Society of America* 16: 2011–21.

Gabriel D., Sait S.M., Hodgson J., Schmutz U., Kunin W.E., Benton T.G. (2010) Scale matters: the impact of organic farming on biodiversity at different spatial scales. *Ecology Letters* 13: 858–69.

Garmenstani A.S., Allen C.R., Gunderson L.H. (2009) Panarchy: discontinuities reveal similarities in the dynamic structure of ecological and social systems. *Ecology and Society* 14(1): 15.

Gibson C.C., Ostrom E., Ahn T.K. (2000) The concept of scale and the human dimensions of global change: a survey. *Ecological Economics* 32: 217-239.

Gimona A., Messager P., Occhi M. (2009) CORINE-based landscape indices weakly correlate with plant species richness in a northern European landscape transect. *Landscape Ecology* 24; 53-64.

Gimona A., van der Horst D. (2007) Mapping hotspots of landscape functions; a case study on farmland afforestation in Scotland. *Landscape Ecology* 22(8): 1255-1264.

Goldman R.L., Thompson B.H., Daily G.C. (2007) Institutional incentives for man-aging the landscape: inducing cooperation for the production of ecosystem services. *Ecological Economics* 64: 333–343.

Gottschalk T.K., Diekötter T., Ekschmitt K., Weinmann B., Kuhlmann F., Purtauf T., Dauber J., Wolters V. (2007) Impact of agricultural subsidies on biodiversity at the landscape level. *Landscape Ecology* 22: 643-656.

Greenberg D.H., Robins P.K. (2008) Incorporating nonmarket time into benefit-cost analyses of cosila programs: an application to the self-sufficiency project. *Journal of Public Economics* 92: 766-794.

Gunderson L., Holling C. (2001) *Panarchy: Understanding Transformations in Human and Natural Systems.* Washington DC: Island Press.

Kampmann D., Lüscher A., Konold W. et al (2012) Agri-environment scheme protects diversity of mountain grassland species. *Land Use Policy* 29(3): 569-576.

Keshkamat, S.S., Tsendbazar N.E., Zuidgeest M.H.P., van der Veen A., de Leeuw J. (2012) The environmental impact of not having paved roads in arid regions: an example from Mongolia. *AMBIO* 41: 202–205.

Kleijn D., Berendse F., Smit R., Gilissen N. (2001) Agrienvironment schemes do not effectively protect biodiversity in Dutch agricultural landscapes. *Nature* 413: 723–725.

Kleijn D., Sutherland W.J. (2003) How effective are European agri-environment schemes in conserving and promoting biodiversity? *Journal of Applied Ecology* 40(6): 947-969.

Kleijn D., Baquero R.A., Clough Y., Díaz M., De Esteban J., Fernández F., Gabriel D., Herzog F., Holzschuh A., Jöhl R., Knop E., Kruess A., Marshall E., Steffan-Dewenter I., Tscharntke T., Verhulst J., West T.M., Yela J.L. (2006) Mixed biodiversity benefits of agri-environment schemes in five European countries. *Ecology Letters* 9(3): 243-254.

Koo B., O'Connell P. (2006) An integrated modelling and multi-criteria analysis approach to managing nitrate diffuse pollution: 1. Framework and methodology. *Science of the Total Environment* 359: 1-16.

Lukesch R., Schuh B. (2010) Working paper on approaches for assessing the impacts of the Rural Development Programmes in the context of multiple intervening factors. Findings of a Thematic Working Group established and coordinated by The European Evaluation Network for Rural Development, 34. [online]URL: http://enrd.ec.europa.eu/evaluation/evaluation-methodologies/assessing-the-rdp-impacts/socio-economic-and-environmental/en/socioeconomic-and-environmental_home_en.cfm.

Mattsson B., Cederberg C., Blix L. (2000) Agricultural land use in life cycle analysis (LCA): case studies in three vegetable oil crop. *Journal of Cleaner Production* 8: 283-292.

MEA (2003) *Ecosystems and Human Well-being: A Framework for Assessment*. Washington DC: Island Press.

Metis GmbH and AEIDL (2008) Synthesis of Ex Ante Evaluations of Rural Development Programmes 2007-2013. Final Report 11/12/2008. Commissioned by: European Commission DG Agriculture and Rural Development.

Michalek J. (2012) Counterfactual impact evaluation of EU rural development programmes - Propensity Score Matching methodology applied to selected EU Member States. Volume 2: A regional approach.Publication No. JRC72060, Publications Office of the European Union.

Midmore P., Langstaff L., Lowman S., Vaughan A. (2008) Evaluating Pillar 2 employment impacts: Case study methodology and results for East Wales. In: Proceedings of 12th Congress of the European Association of Agricultural Economists, 26-29 August 2008, Ghent, Belgium.

Montresor E., Pecci F., Pontarollo N. (2010) Rural development policies at regional level in the enlarged EU. The impact on farm structures. Paper prepared for presentation at the 114th EAAE Seminar ‘Structural Change in Agriculture’, Berlin, Germany, April 15 - 16, 2010.

Mortimer S.R., Mauchline A.L., Park J.R., Finn J.A., Edwards D., Morris J. (2010) Evaluation of agri-environment and forestry schemes with multiple objectives. *EuroChoices* 9: 48-54.

Paracchini M.L., Britz W. (2010) Quantifying effects of changed farm practices on biodiversity in policy impact assessment – an application of CAPRI-Spat, Ispra, Institute for Environment and Sustainability of the Joint Research Centre, European Commission.

Pecci F., Sassi M. (2008) A mixed geographically weighted approach to decoupling and rural development in the EU-15. Paper prepared for presentation at the 107th EAAE Seminar "Modelling of Agricultural and Rural Development Policies". Sevilla, Spain, January 29th - February 1st, 2008.

Pelosi C., Goulard M., Balent G. (2010) The spatial scale mismatch between ecological processes and agricultural management: Do difficulties come from underlying theoretical frameworks? *Agriculture, Ecosystems & Environment* 139: 455–462.

Prager K., Reed M., Scott A. (2012) Encouraging collaboration for the provision of ecosystem services at a landscape scale – rethinking agri-environmental payments. *Land Use Policy* 29: 244-249.

Primdahl J., Vesterager J.P., Finn J., Vlahos G., Kristensen L., Vejre H. (2010) Current use of impact models for agri-environment schemes and potential for improvements of policy design and assessment. *Journal of Environmental Management* 91: 1245–54.

Primdahl J., Peco B., Schramek J., Andersen E., Oñate J.J. (2003) Environmental effects of agri-environmental schemes in Western Europe. *Journal of Environmental Management* 67(2): 129-138.

Princé K., Moussus J., Jiguet F. (2012) Mixed effectiveness of French agri-environment schemes for nationwide farmland bird conservation. *Agriculture, Ecosystems & Environment* 149: 74-79.

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

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

Reidsma P., Ewert F. (2008) Regional farm diversity can reduce vulnerability of food production to climate change. *Ecology and Society* 13(1): 38. [online] http://www.ecologyandsociety.org/vol13/iss1/art38/

Reidsma P., Ewert F., Oude Lansink A., Leemans R. (2009) Vulnerability and adaptation of European farmers: a multi-level analysis of yield and income responses to climate variability. *Regional Environmental Change* 9: 25-40.

Reinhard A. J., Linderhof V. (2013) Using spatial econometrics in impact assessment. SPARD deliverable D4.5. The Hague, LEI-Wageningen UR.

Reinhard S., Linderhof V., van Leeuwen E., Smit M.J., Nowicki P., Michels R. (2013). Spatial econometric models for evaluating RDP measures: analyses for the EU27 SPARD (Vol. SPARD deliverable D4.3). Den Haag (the Netherlands): LEI-Wageningen UR.

Righi E., Dogliotti S., Stefanini F.M., Pacini G.C. (2011) Capturing farm diversity at regional level to up-scale farm level impact assessment of sustainable development options. *Agriculture, Ecosystems & Environment* 142: 63–74.

Roth D., Schwabe M. (2003) Method for assessing the proportion of ecologically, culturally and provincially significant areas (OELF) in agrarian spaces used as a criterion for environmental friendly agriculture. *Agriculture, Ecosystems & Environment* 98: 435–441.

Saez C.A., Requena J.C. (2007) Reconciling sustainability and discounting in cost-benefit analysis: a methodological proposal. *Ecological Economics* 60: 712-725.

Schneider D. (2001) The rise of the concept of scale in ecology. BioScience 51(7): 545-553.

Sheppard S.J. (2005) Landscape visualisation and climate change: the potential for influencing perceptions and behaviour. *Environmental Science & Policy* 8(6): 637-654.

Steinhardt U., Volk M. (2003) Mesoscale landscape analysis on the base of investigations of water balance and water-bound material fluxes: Problems and hierarchical approaches for their resolution. *Ecological Modelling* 168: 251-265.

Stoeglehner G., Narodoslawsky M. (2008) Implementing ecological footprinting in decision-making processes. *Land Use Policy* 25(3): 421-431.

Stolze M., Piorr A., Häring A.M., Dabbert S. (2000) Environmental impacts of organic farming in Europe. *Organic Farming in Europe: Economics and Policy*, Vol. 6. Universität Hohenheim, Stuttgart-Hohenheim.

Teillard F., Allaire G., Cahuzac E., Léger F., Maigné E., Tichit M. (2012) A novel method for mapping agricultural intensity reveals its spatial aggregation: Implications for conservation policies. *Agriculture, Ecosystems & Environment* 149: 135–143.

Terluin I., Berkhout, P (2011) Exploring the perspectives of a mixed case study approach for the evaluation of the EU Rural Development Policy 2007-2013. Paper prepared for the 122nd

EAAE Seminar "Evidence-Based Agricultural and Rural Policy Making: Methodological and Empirical Challenges of Policy Evaluation" Ancona, February 17-18, 2011.

Terluin I., Roza P. (2010) Evaluation methods for rural development policy, Report 2012-037, LEI, Wageningen UR999999.

Van den Bergh J.C.J.M., Verbruggen H. (1999) Spatial sustainability, trade and indicators: an evaluation of the ecological footprint. *Ecological Economics* 29: 61-72.

Van der Horst D. (2007) Assessing the efficiency gains of improved spatial targeting of policy interventions; the example of an agri-environmental scheme. *Journal of Environmental Management* 85: 1076-1087.

Van Ittersum M.K., Brouwer F. (2010) Introduction. In: Brouwer F., Van Ittersum M.K. (Eds.) *Environmental and Agricultural Modelling: Integrated Approaches for Policy Impact Assessment*. Springer, Dordrecht, pp 1-7.

Van Ittersum M.K., Ewert F., Heckelei T., Wery J., Alkan Olsson J., Andersen E., Bezlepkina I., Brouwer F., Donatelli M., Flichman G., Olsson L., Rizzoli A.E., van der Wal T., Wien J.E., Wolf J. (2008) Integrated assessment of agricultural systems - A component-based framework for the European Union (SEAMLESS). *Agricultural Systems* 96: 150-165.

Veldkamp T., Polman N., Reinhard S., Slingerland M. (2011) From scaling to governance of the land system: Bridging ecological and economic perspectives. *Ecology and Society* 16(1).

Vermaat J.E., Eppink F., van den Bergh J.C.J.M., Barendregt A., van Belle J. (2005) Aggregation and the matching of scales in spatial economics and landscape ecology: empirical evidence and prospects for integration. *Ecological Economics* 52(2): 229-237.

Vervoort J.M., Rutting L., Kok K., Hermans F.L.P., Veldkamp T., Bregt A.K., van Lammeren R. (2012) Exploring dimensions, scales and cross-scale dynamics from the perspectives of change agents in social-ecological systems. Ecology and Society 17(4): 24.

Whittingham M.J. (2011) The future of agri-environmental schemes: biodiversity gains and ecosystem service delivery? *Journal of Applied Ecology* 48: 509-513.

Willemen L., Hein L., van Mensvoort M.E.F., Verburg P. H. (2010) Space for people, plants, and livestock? Quantifying interactions among multiple landscape functions in a Dutch rural region. *Ecological Indicators* 19: 62-73

Willemen L., Verburg P.H., Hein L., van Mensvoort M. E. F. (2008) Spatial characterization of landscape functions. *Landscape and Urban Planning* 88: 34-43.

Williams K.J., Reeson A.F., Drielsma M.J., Love J. (2012) Optimised whole-landscape ecological metrics for effective delivery of connectivity-focused conservation incentive payments. *Ecological Economics* 81: 48–59.

Zucker J. (2006) Analyse der Leistungsfähigkeit und des Nutzens von Evaluationen der Politik zur Entwicklung ländlicher Räume in Deutschland und Großbritannien am Beispiel der einzelbetrieblichen Investitionsförderung. Landbauforschung Völkenrode: Sonderheft 293. Braunschweig: Bundesforschungsanstalt für Landwirtschaft.

Table A. 1 Summary of models from literature

Author(s)	*Year*	Methodology	Objective	RDP /public good
Amano et al	2011	Multilevel nested regressions (hierarchical linear modelling) and MCMC simulation.	An assessment of the macro-level factors (climate and topography) on the effect of farming practices on biodiversity.	Biodiversity
Araujo et al	2005	Generalised additive modelling (GAM)	Downscaling European species atlas distribution to a finer resolution	Biodiversity
Aviron et al	2007	Mixed general linear models (GLM), multivariate analysis, redundancy analysis and MCMC simulation for two scales, local and landscape	Assessment of the impact of AES on butterfly biodiversity in the context of two scales: local grassland conditions and the landscape context of the grasslands.	Biodiversity
Baker et al.	2012	Statistical modelling approach (multivariate loglinear model) for three different spatial resolutions ($1km^2$, $9\ km^2$ and $25km^2$)	Assessment of the effect of AES on bird species that use agricultural land during their life-cycle and are expected to benefit from AES management.	Biodiversity
Carey et al.	2003	Expert panel assessment of ecological, landscape, historical and access objectives of 484 CSS agreements	To develop a multi-disciplinary approach to assess the degree to which ecological, landscape historical and access objecitves of AES are met by looking at CSS only	Landscape, biodiversity
Carey et al.	2005	Expert panel assessment of individual AES agreements in landscape context	To develop a multi-disciplinary approach to assess the degree to which ecological, landscape historical and access objecitves of AES are met (CSS and ESA)	Landscape, biodiversity, cultural services
Castellazzi et al.	2010	Simulation scenarios using the LandSFACTS model		

Author(s)	*Year*	Methodology	Objective	RDP /public good
Christie and Gibbons	2011	Choice experiment models, Bayes	Develop methodology to estimate and account for respondents' ability to choose (ATC)	Biodiversity, coastal defences
Concepcion et al.	2012, 2011, 2008	Landscape metrics (size of focal field, length of boundaries around field, proportion of areas occupied by non-productive land use (within a radius of 500m), and generalised mixed effect models (GMMs),	An analysis of the impact of field management intensity on biodiversity along a wide gradient of landscape complexity. Data on species richness of birds, plants, spiders and bees were used in 232 extensively and intensively managed paired fields (112 arable fields and 120 grasslands) in 18 regions distributed across six European countries	Biodiversity
Courtney et al.	2013	Interviews with ES agreement holders, plus adapted LM3 model	To capture the direct, indirect and induced effects of a variety of ES schemes at sub-regional level	N/A
Davey et al.	2010	Generalised Linear Models	Assessment of the evidence for impact of Entry Level Stewardship on biodiversity conservation in the wider countryside based on measured farmland bird abundances	Biodiversity
De Benedetto and Klemeš	2009	Life Cycle Assessment	Use LCA to produce an Environmental performance Strategy Map – a graphical map which combines the main environmental indicators with cost	Water, carbon footprint, energy
Deconchat et al.	2007	Multidisciplinary nested sampling approach	Social and Ecological Assessment of the impact of forest and agricultural management practices on biodiversity	Biodiversity
Delattre et a.l	2013	Monitoring	Monitor behaviour of butterflies to test the use of GFMs as ecological corridor networks	Biodiversity
Ewert et al.	2011	Integration framework based on a set of methods and approaches for an integrated assessment (SEAMLESS-IF)	A conceptual analysis of scale changes and methods of model integration for addressing complex integrated assessment problems in agri-environmental systems, aimed at an integrated assessment of policy impacts on agricultural systems.	

Author(s)	*Year*	Methodology	Objective	RDP /public good
Gabriel et al.	2010	Multi-scale hierarchical sampling programme to estimate species density	Disentangle the scale effects of farming management on biodiversity and identify those scales most appropriate for maximizing the biodiversity benefits in agricultural landscapes.	Biodiversity
Gimona and van der Horst	2007	Landscape zoning based on multiple criteria	Evaluation of actual AES activities/output against potential benefit maps of public goods (biodiversity, visual amenity and recreation), assessment of the efficient of a specific AE scheme in delivering public goods.	Biodiversity, landscape and recreation
Gottschalk et al.	2007	Model based on GLM	Examining the long-term effects of two AES on the species' richnesss of birds and carabids in marginal regions of Europe	
Kampmann et al.	2012	Principle component analysis	Assessment of the effectiveness of AES and the variability of this effectiveness under different climate and socio-economic conditions	
Kleijn et al.	2006	Field pairing, measurement of species density and abundance	Evaluation of the biodiversity effects of conservation management on farms across a wide range of EU agricultural landscapes	Biodiversity
Koo and O'Connell	2006	Multi Criteria Analysis	Develop a land optimisation methodology as a compromise between long-term nitrate pollution and agronomy at the catchment scale	Water quality
Primdahl et al.	2003	Interviews among beneficiaries	Assessment of environmental effects of AE schemes through an assessment of agricultural practices and agreements.	Water & soil quality
Princé et al.	2012	Markov Chain Monte Carlo using breeding bird survey	To evaluate the effectiveness of French AES in enhancing farmland bird diversity at a national scale	Biodiversity
Reidsma and	2008	Statistical modelling approach	Impact assessment farm biodiversity on farm vulnerability to	Climate change

Author(s)	*Year*	Methodology	Objective	RDP /public good
Ewert			climate change.	
Reidsma et al.	2009	Statistical modelling approach	Multi-level analysis of yield and income responses to climate variability	Climate change
Reinhard et al.	2013	Spatial econometrics	Develop a modelling tool to understand the causal relationship between rural development measures and their results in a spatial dimension	Biodiversity, water quality, tourism
Righi et al.	2011	Cluster analysis (CA), multidimensional scaling (MDS) and similarity percentages (SIMPER) analysis	1. Identify farm typologies 2. Integrate the typology into a quantitative system approach to upscale farm-level results	Soil
Roth and Schwabe	2003	Calculation of target values for OELF per natural region/landscape character	To assess the proportion of ecological, cultural and regional valuable areas (OELF) through spatially specific current values and targets	All public goods
Teillard et al	2012	Hierarchical clustering analysis	An analysis of the spatial distribution of agricultural intensity, to examine the distribution of agricultural production types against the intensity gradient and the implications of spatial aggregation of intensity for conservation policy	Biodiversity (HNV)
Willemen et al	2008	Landscape function quantified through one of three different methods of delineation using either existing data sources of observations.	Development of methodology for the quantification of landscape functions and visualising their spatial variability	Landscape, Agriculture, HNV, Water
Willemen et al	2010	A three step approach: 1) quantification of landscape functions (Willemen et al., 2008), 2)	Evaluation of the impact of regional development policies on future landscape services	All Public goods

Author(s)	*Year*	Methodology	Objective	RDP /public good
		landscape service supply through thresholds, and 3) is landscape service value through monetary valuation.		
Williams et al	2012	Spatial modelling using landscape ecological metrics	Assessment of effective delivery of connectivity-focused conservation measures	Biodiversity (species conservation)