



European Evaluation Network
for Rural Development

Draft Working Paper on:
Approaches for assessing the impacts
of the rural development programmes in the
context of multiple intervening factors

26 November 2009

Editors: R. Lukesch
B. Schuh

Contributions: G. Beaufoy (HNV)
H. Goemann (Water quality)
P. Kaufmann (Employment creation)
P. Koorberg (Biodiversity)
J. Michalek (Economic growth)
D. Moran (Climate change)
M. L. Paracchini (HNV)
G. Pinay (Water quality)
A. Pufahl (Employment creation)
S. Schiller (Labour productivity)
P. Rossi (Biodiversity)
D. Storti (Labour productivity)

DRAFT

CONTENTS

1.	Introduction	7
1.1	About this Guidance Document	7
1.1.1	Rationale	7
1.1.2	Purpose	7
1.1.3	Genesis	8
1.1.4	Limitations	9
1.2	Structure of the Guidance Document	9
2.	The CMEF requirements for the assessment of RDP impacts	11
2.1	The importance of assessing impacts	11
2.2	The intervention logic	12
2.3	Main elements	13
2.3.1	The common impact indicators	13
2.3.2	Trends in baseline indicators	13
2.3.3	Additional programme specific indicators	14
2.3.4	Common and specific evaluation questions	14
3.	The core process of assessing RDP impacts	16
3.1	The three-phase process of assessment of impacts	16
3.1.1	First phase: Gauging the evidence of change	16
3.1.2	Second phase: Identifying the drivers of change	17
3.1.3	Third phase: Understanding change and concluding on future interventions	19
3.2	The Evaluation Architecture	20
3.2.1	Dealing with uncertainties	20
3.2.2	Reducing complexity through a consistent approach	20
3.2.3	Constraints in utilization of the evaluation results	23
3.3	Assessing impacts of RDP – challenges and practical solutions	23
3.3.1	Methodological challenges	23
3.3.2	Five key issues	25
3.3.3	The counterfactual assessment of impacts	25
3.3.4	Taking into account and cross-relating impacts at micro and macro level	33
3.3.5	Netting out the programme effects by reducing deadweight, leverage, displacement, substitution and multiplier effects	39
3.3.6	Notes on data collection and processing	45
3.3.7	Bridging the gap between measuring impact indicators and providing answers on programme impacts	48
4.	Guidance for assessing the socio-economic impact indicators and addressing the related evaluation questions	53
4.1	Impact Indicator 1: Economic Growth	53
4.1.1	The CMEF requirements	53
4.1.2	Key challenges with regard to measurement and interpretation	55
4.1.3	Recommendable methods of measurement	57
4.1.4	Data requirements and collection	60
4.1.5	Interpretation and judgement issues	61
4.2	Impact Indicator 2: Employment Creation	62
4.2.1	The CMEF Guidelines	62
4.2.2	Key challenges with regard to measurement and interpretation	63
4.2.3	Recommendable methods of measurement	64

4.2.4	Data requirements and collection	71
4.2.5	Interpretation and judgment issues	73
4.3	Impact Indicator 3: Labour Productivity	75
4.3.1	The CMEF guidelines	75
4.3.2	Key challenges with regard to measurement and interpretation	76
4.3.3	Recommendable methods of measurement	76
4.3.4	Data requirements and collection	85
4.3.5	Interpretation and judgement issues	86
5.	The assessment of environmental RDP impacts	89
5.1	Impact Indicator 4: Reversing Biodiversity Decline	90
5.1.1	The CMEF guidelines	90
5.1.2	Key challenges with regard to measurement and interpretation	93
5.1.3	Recommendable methods of measurement	94
5.1.4	Data requirements and collection	99
5.1.5	Interpretation	104
5.1.6	Judgement issues	109
5.2	Impact Indicator 5: Maintenance of HNV farming and forestry	112
5.2.1	The CMEF guidelines	112
5.2.2	Key challenges with regard to measurement and interpretation	115
5.2.3	Recommendable methods of measurement	128
5.2.4	Data requirements and collection	130
5.2.5	Interpretation	130
5.2.6	Judgment issues	132
5.3	Impact Indicator 6: Improvement in Water Quality	134
5.3.1	The CMEF guidelines	134
5.3.2	Key challenges with regard to measurement and interpretation	135
5.3.3	Recommendable methods of measurement	135
5.3.4	Data requirements and collection	139
5.3.5	Interpretation	141
5.3.6	Judgment issues	143
5.4	Impact Indicator 7: Contribution to Combating Climate Change	146
5.4.1	The CMEF guidelines	146
5.4.2	Related Common Evaluation Questions and Horizontal Evaluation Questions according to Guidance Note B	148
5.4.3	Key challenges with regard to measurement and interpretation	150
5.4.4	Recommended methods of measurement	152
5.4.5	Data requirements and collection	156
5.4.6	Judgment issues	158
6.	Bibliography	160
7.	Abbreviations	167

Tables and Figures

Table 1	Rural development programme interventions and impact indicators	18
Table 2	Availability of data by source and geographical level	46
Table 3	Six methods for regional environmental impact assessment (from Payraudeau & van der Werf 2005)	139
Figure 1	Relations between the baseline indicators, output, result and impact indicators	12
Figure 2	Causal loops connect different challenges	38
Figure 3	TERESA regional system dynamics model	39
Figure 4	Approach to establish gross and net programme effects at micro-level	76
Figure 5	An example of single species index trend in Europe. <i>Alauda arvensis</i> (Skylark) is a typical farmland bird species.	94
Figure 6	An example of aggregated indices trends	96
Figure 7	The Three Key Characteristics of HNV Farming Systems	113
Figure 8	Interactions of models for the Weser River basin.	142
Figure 9	Schematic presentation of MITERRA-EUROPE	145

DRAFT

1. Introduction

1.1 About this Guidance Document

1.1.1 Rationale

Council Regulation 1698/2005 on support for rural development by the European Agricultural Fund for Rural Development – EAFRD provides the legal framework for the preparation and implementation of rural development programmes in the Member States for the period 2007 – 2013. Art. 80 of Regulation 1698/2005 stipulates a Common Monitoring and Evaluation Framework (CMEF) which has been drawn up in cooperation between the Commission and the Member States.

The new arrangements provide a single framework for monitoring and evaluation of all rural development interventions. It provides broad continuity as regards monitoring requirements and constitutes a significant simplification as regards assessment of results and impacts, while at the same time offering greater flexibility to Member States.¹

The CMEF Handbook, consisting of a Guidance Document and four annexes (comprising 15 chapters) embodies the effort to continue and adapt the guidance for the 2000-2006 period on the basis of experience and the requirements of the new regulation; integrating the state of the arts of evaluation on a global scale. The assessment of impacts is of particular importance for justifying policy interventions, and at the same time it poses considerable conceptual and methodological challenges.²

The Member States have reported difficulties in identifying the impacts attributable to specific measures in the context of multiple intervening factors. The challenges regarding the establishment of the intervention logic are particularly great in the case of the environmental impacts (common impact indicators 4-7), because these impacts are strongly influenced by site-specific circumstances (e.g. soil, temperature, rainfall). Moreover, both environmental and socio-economic impacts (reflected by the common impact indicators 1-3) may take a long time to emerge and may depend on other intervening factors (e.g. national/regional policies, implementation mechanisms).

1.1.2 Purpose

This Guidance Document serves to cover these aspects. The document will allow for a pragmatic handling of measure-specific impacts, and includes – where applicable – relevant guidance for the establishment and analysis of the counterfactual. Rather than determining a common EU method for addressing these issues, the Guidance paper suggests a set of approaches, in order to allow Member States to capitalize on the work they have already undertaken on these topics. With this in mind, attention is given to collecting and utilizing examples of current practice across the EU. This Guidance Paper aims at operationalising the Common Monitoring and Evaluation Framework (CMEF) with regard to the assessment of impacts. The document responds to the requirement – as listed in the Common Monitoring and Evaluation Framework (Guidance Note A): “*The Methodology for the estimation of impact*

¹ Handbook on Common Monitoring and Evaluation Framework – Guidance Document. Sep. 2006, DG Agri, Bruxelles, p.5f.

² As a cross-reference to the state of the arts of impact assessment as shared by the global community of evaluators, see the “NONIE Guidance on Impact Evaluation” (Leeuw F., Vaessen J. 2009): <http://www.worldbank.org/ieg/nonie/guidance.html>

will be developed further during the programme period by the Evaluation Network.” – meaning that this document describes the possible methodologies to be applied for assessment of impacts in the CMEF context. Therefore its purpose is:

- ▶ to support the responsible administrations in Member States to position their approaches in the state of the art in the evaluation community and in selecting appropriate evaluation methods and tools;
- ▶ to develop convergent viewpoints on assessment of impacts for rural development programmes across Europe.

1.1.3 Genesis

A Thematic Working Group of the European Evaluation Network for Rural Development, consisting of two sub-groups – for socio-economic and environmental indicators respectively – has been set up in order to examine viable approaches to assessing the impacts of selected rural development programmes in the context of multiple intervening factors. The Guidance paper is the result of this work in which evaluation experts and practitioners were involved during 2009.

To this end, the Helpdesk and involved experts

- ▶ have examined the methodological state of the art of assessment of impact on a global scale;
- ▶ have conducted a survey among Member States for assessing their specific difficulties and challenges in respect to impact assessment (e.g. for which measure, or methodological questions);
- ▶ have asked Member States and the related network of evaluators to come up with their practices of evaluation of impacts in order to make them better known to the whole community as possibly illuminating examples for others to meet their challenges;
- ▶ have synthesised the findings in this guidance document, which provides general support for the assessment of impacts of rural development programmes as well specific support for assessing the seven common impact indicators.

1.1.4 Limitations

What could not be done although it was intended in the planning phase of this work³ was

- ▶ widening the systemic focus of the evaluation system as currently presented; hence this paper does not go into the matter of assessing alternative indicators and measurement methods of specific measures;
- ▶ compiling a collection of demonstrable practices to modelling cause-effect chains and dealing with the attribution gap, which go beyond the quotation of and reference to current practices in Member States.

The reason for these limitations is that already in the survey among Member States it became clear that the main issue would be to explain the CMEF, make it more operational and better understood, in order to achieve sound, compatible and comparable evaluation designs across European Member States and regions. The Member States were mostly setting up their evaluation systems during 2009 so that good practice examples could almost exclusively be provided from the last period 2000-2006. The ongoing work of the Helpdesk will guarantee an updating of good practices in operation during this period, in order to foster exchange, dissemination and discussion among the concerned authorities and experts in the Member States and in the European Commission.

1.2 Structure of the Guidance Document

The paper is divided into six sections. The **first section** provides an overview on the purpose and the structure of this document. The **second section** recalls and explains the main requirements for assessment of impacts as laid down in the Common Monitoring and Evaluation Framework for Rural Development Programmes⁴, integral part of the Council Regulation 1698/2005 relating to the European Agricultural Fund for Rural Development (EAFRD). The **third section** elaborates on the main challenges in assessing impacts and provides an overview on the three main phases of the evaluation architecture.

On the one hand it seems clear that socio-economic and environmental impacts are inseparable while evaluating the overall impact of an RD programme. However, the respective methodological requirements for assessing these impacts (as addressed by the seven common impact indicators) suggest their separate discussion in this working document:

- ▶ The assessment of socio-economic and environmental impacts refers to methodologies which have their background in different scientific realms (social versus natural sciences);
- ▶ The similarities of approaches are considerable in respect to the three common socio-economic impact indicators (economic growth, employment creation and labour productivity), less so but still tangible concerning the four common environmental impact indicators (biodiversity decline, HNV farming and forestry, water quality and combating climate change).

³ Helpdesk of the European Evaluation Network for Rural Development (2009): Note regarding the organisation of activities 1.1.1 and 1.1.2 of the Annual Work Programme 2009 as a joint thematic work group: "Approaches for assessing the impacts of the rural development programmes in the context of multiple intervening factors." Bruxelles.

⁴ http://ec.europa.eu/agriculture/rurdev/eval/index_en.htm

The guidance document therefore focuses on assessing the socio-economic impacts in the **fourth section** and discusses the assessment of the environmental impacts in the **fifth section**. Each of these sections which make up most of this document is introduced with a sub-chapter discussing methodological challenges concerning general aspects of

- ▶ assessing socio-economic impacts and the related three indicators (economic growth, employment creation and labour productivity), and
- ▶ assessing environmental impacts and the related four indicators (biodiversity decline, HNV farming and forestry, water quality and combating climate change).

The chapters four and five are all structured in the same way:

- ▶ The first sub-chapter recalls the CMEF requirements for the specific indicator;
- ▶ the second sub-chapter sets out the key challenges;
- ▶ the third sub-chapter describes recommendable methods of measurement;
- ▶ the fourth sub-chapter focuses on requirements for data and their collection;
- ▶ the fifth sub-chapter discusses the possibilities and limits of interpretation and judgement.

The **sixth section** lists the bibliographic sources, documents and websites consulted and used for the production of this Guidance Document.

2. The CMEF requirements for the assessment of RDP impacts

2.1 The importance of assessing impacts

The assessment of impacts of a given policy intervention is important for the following reasons:

- ▶ It provides empirical evidence on whether a specific policy worked or did not work. It also provides an information about the sustainability of effects of a given policy intervention;
- ▶ By comparing results of a policy intervention with target values it provides an information on effectiveness of a given policy intervention and achievability of more general societal goals (e.g. concerning growth or development) using this specific policy instrument;
- ▶ It helps to re-design a policy intervention (programme) to make it more effective and efficient (by taking into consideration costs of intervention);
- ▶ It provides arguments for continuation or discontinuation of policies/programmes by comparing social benefits with costs of specific policy interventions;
- ▶ It helps to learn about functioning of economic, social and environmental processes;
- ▶ It improves institutional capacities of organisations involved in impact evaluations;
- ▶ It improves decision making at all levels;
- ▶ It provides some information regarding accountability of institutions involved in formulation and implementation of policies.

In order to foster a common perspective and approach to impact evaluation of rural development programmes and their potential impacts on people, society and the environment, the European Commission has produced a Common Monitoring and Evaluation Framework (CMEF) Handbook⁵ defines impacts as

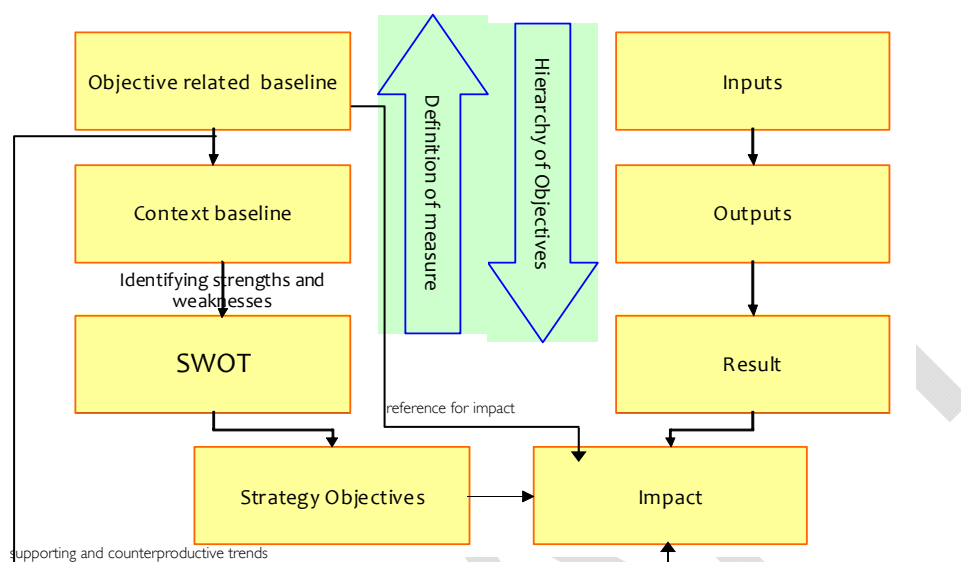
effects of an intervention lasting in medium or long term. Some impacts appear indirectly (e.g. turnover generated for the suppliers of assisted firms). Some can be observed at the macro-economic or macro-social level (e.g. improvement of the image of the assisted area); these are global impacts. Impacts may be positive or negative, expected or unexpected.

⁵ Handbook on Common Monitoring and Evaluation Framework – Guidance note N: Glossary of terms; DG Agriculture (2007).

2.2 The intervention logic

The backbone of the CMEF is the intervention logic of RD programmes, linking inputs, outputs, results and impacts and relating these to the programme objectives in the way as depicted here:

Figure 1 Relations between the baseline indicators, output, result and impact indicators



Source: Handbook on Common Monitoring and Evaluation Guidelines – Guidance Document p.7 (own additions)

According to the intervention logic of rural development programmes laid down in the CMEF Handbook, impacts represent the final link of the results chain which starts with the input (intervention), producing an output whose use by the beneficiaries brings forth results, which in turn contribute to the impact. Impacts are therefore correlated with the three core objectives of the rural development programmes 2007-2013:

- ▶ Improving the competitiveness of agriculture and forestry by support for restructuring, development and innovation;
- ▶ Improving the environment and the countryside by supporting land management;
- ▶ Improving the quality of life in rural areas and encouraging diversification of economic activity;

The evidence for impacts shall be provided by indicators which⁶

refer to the benefits of the programme beyond the immediate effects in its direct beneficiaries both at the level of the intervention but also more generally in the programme area. They are linked to the wider objectives of the programme. They are normally expressed in “net” terms, which means subtracting effects that cannot be attributed to the intervention (e.g. double counting, deadweight), and taking into account indirect effects (displacement and multipliers). Example: increase in employment in rural areas, increased productivity of the agricultural sector, increased production of renewable energy.

⁶ ibidem

2.3 Main elements

According to the CMEF, all the elements presented in this sub-chapter should be taken into consideration when assessing the impact of RD programmes.

2.3.1 The common impact indicators

Following this logic assessment of impacts first of all, but not exclusively, relies on the seven common impact indicators:

The Common Monitoring and Evaluation Framework (CMEF) foresees seven common impact indicators relating to growth, jobs, productivity, biodiversity, high nature value areas, water and climate change which reflect explicitly objectives established by the European Council and the Strategic Guidelines for rural development. The impact of the programme as a whole should be assessed against these seven indicators to take into account the full contribution of all axes of the programme.⁷

2.3.2 Trends in baseline indicators

In addition, assessment of impacts will be supported by establishing impacts through the comparison of the baseline trends – by using the set of common baseline indicators and the change of indicator values throughout the life cycle of the RD programme:

Objective related baseline indicators: These are directly linked to the wider objectives of the programme. They are used to develop the SWOT analysis in relation to objectives identified in the regulation. They are also used as a baseline (or reference) against which the programmes' impact will be assessed. Baseline indicators reflect the situation at the beginning of the programming period and a trend over time. The estimation of impact should reflect that part of the change over time that can be attributed to the programme once the baseline trend and other intervening factors have been taken into account.

Context related baseline indicators: These provide information on relevant aspects of the general contextual trends that are likely to have an influence on the performance of the programme. The context baseline indicators therefore serve two purposes: (i) contributing to identification of strengths and weaknesses within the region and (ii) helping to interpret impacts achieved within the programme in light of the general economic, social, structural or environmental trends.⁸

However – *In many cases, it will not be possible to link directly programme impacts with baseline trends due to the scale of the intervention or the lack of baseline data at an appropriate level. Evaluation of impact should therefore focus on a bottom-up approach to assessing programme effects. Evaluators should seek to assess the link between the impact of the programme and baseline trends, but this does not necessarily need to be quantified.⁹*

⁷ Handbook on Common Monitoring and Evaluation Framework – Guidance Note A; DG Agriculture (2007)

⁸ Handbook on Common Monitoring and Evaluation Framework, "Monitoring and Evaluation 2007-2013, page 8; DG Agriculture (2007)

⁹ Handbook on Common Monitoring and Evaluation Framework – Guidance Note A; DG Agriculture (2007)

2.3.3 Additional programme specific indicators

Since common indicators may not fully capture all effects of programme activity, in particular for national priorities and site-specific measures, it is necessary to define additional indicators within the programmes¹⁰. Such additional indicators should be developed by Member States and programme partnerships in a flexible manner, but in accordance with the general principles governing the use of indicators in the CMEF. There are a number of situations where Member States should provide additional indicators:

- ▶ When a common baseline indicator does not cover the specific characteristics of the programme area.
- ▶ When an additional objective or national priority defined in the National Strategy Plan or the programme is not covered by an impact indicator;
- ▶ When common impact indicators are not detailed or specific enough to reflect the wider benefits of a measure, or where a common impact indicator does not exist for a measure. This is particularly important where measures are highly site-specific, for example in agri-environment. Appropriate measure-specific impact indicators should be developed.
- ▶ When common result indicators are not detailed or specific enough to reflect the first effects of a measure, or where a common result indicator does not exist for a measure.
- ▶ When common output indicators are not detailed or specific enough to reflect the activities under a measure;

The definition of additional programme specific indicators will give Member States flexibility in creating a monitoring and evaluation system adapted to their needs. Nevertheless this flexibility is only possible as long as it stays within the scope of the rural development regulation and the corresponding hierarchy of objectives. In developing additional indicators Member States:

- ▶ Ensure the relevance and utility of an additional indicator;
- ▶ Define the type and use of the indicator;
- ▶ Ensure that the additional indicator meets accepted quality criteria for the type of indicator and intervention concerned.

It is recommended that a detailed indicator fiche for each additional indicator is provided to facilitate their use in monitoring and evaluation.¹¹

2.3.4 Common and specific evaluation questions

As indicator measurement is a means and not the end of assessment of impacts, their specific function lies in the evidence they should provide for answering the Common Evaluation Questions laid down in the CMEF (i) for each axis and (ii) across all axes¹². The answers to these questions should be underpinned by the evidence based on indicator measurement or estimation and the relevant

¹⁰ Additional indicators are sometimes referred to as programme specific indicators. They are the same.

¹¹ Handbook on Common Monitoring and Evaluation Framework – Guidance Note A; DG Agriculture (2007). Please note that the CMEF uses the terms “additional indicator”, “additional programme specific indicator” synonymously and for all indicator types – i.e. baseline, output, result and impact.

¹² in which case they are called *Horizontal Evaluation Questions*

judgement criteria, as well as by additional quantitative and qualitative information from public statistics, specific surveys/enquiries, or other sources¹³.

For the sake of gaining relevance in assessment of impacts of a specific programme, the common evaluation questions should be complemented by programme-specific evaluation questions¹⁴.

For setting up the evaluation system, it is inevitable for the managing authority to review the common and programme specific evaluation questions and the related indicators in order to assess what needs to be done in terms of information gathering and analysis in order to answer these questions in a meaningful and appropriate manner.

Whereas the common evaluation questions and indicators are defined in a manner that makes them applicable across a large number of programmes, more precise target levels reflected by indicators and more precise questions may need to be established by the programming authority.

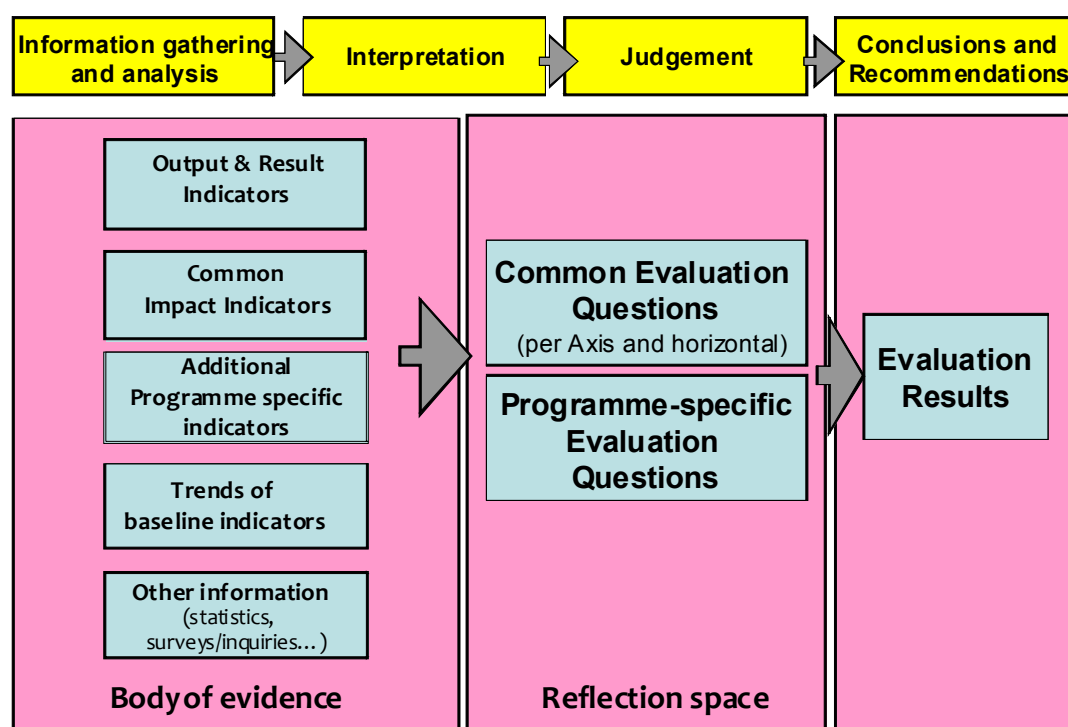
¹³ Handbook on Common Monitoring and Evaluation Framework – Guidance Note B: Evaluation Guidelines; DG Agriculture (2007)

¹⁴ ibidem

3. The core process of assessing RDP impacts

3.1 The three-phase process of assessment of impacts

The process of assessment of impacts of Rural Development programmes consists of three distinctive phases which are depicted in the following chart¹⁵.



3.1.1 First phase: Gauging the evidence of change

This phase provides an answer to the question if the rural development programme has brought forth measurable changes which do not occur or would not have occurred due to other influences such as general trends, external economic shocks, the impact of other policy interventions etc. Building up this “body of evidence” includes the structuring of the evaluation process, data collection and information gathering, the measurement of changes in the indicator values and their interpretation. Evidence for the occurrence and extent of change induced by policy interventions is essential for justifying public spending in the respective policy field.

The responsibility of the Managing Authority is to ensure that the evaluators have sufficient data on general trends, outputs and results to carry out such an assessment. The responsibility for the estimation and quantification of impact remains with independent evaluators. Evaluators will often find it necessary to

¹⁵ “Information gathering and analysis” involves the accomplishment of the tasks “structuring”, “observing” and “analysing” as stipulated in the CMEF: Handbook on Common Monitoring and Evaluation Framework – Guidance Note B: Evaluation Guidelines; DG Agriculture (2007)

undertake further investigation and to add elements of qualitative analysis to estimate impact and interpret baseline data in the specific context of the beneficiaries of the programme.

In general, impact indicators should be expressed (i) in absolute amounts (to estimate cost-effectiveness) and (ii) in relation to those beneficiaries affected by the intervention directly and, where appropriate, to those affected indirectly. There is a specific focus on quantification of impact in the rural development regulation, particularly in relation to the baseline situation. However, in many cases, although it is possible to assess the baseline situation and impact at the level of the direct and indirect beneficiaries of the support, it is often more difficult to place this in the context of the more general baseline trends at the level of the programme area. This may be linked to as highlighted above to the relatively small scale of the intervention or lack of appropriate baseline data. For this reason, the focus should be on the bottom-up estimation of impact:

- ▶ In a first step, impact should be estimated at the level of direct and indirect beneficiaries by programme evaluator on the basis of output and result indicators, survey data and benchmark data and coefficients from similar projects and past experience and evaluations (for calculation of double counting, deadweight, displacement and multiplier effects). This should be cross-checked against the counterfactual situation and contextual trends in the programme area.
- ▶ In a second step, the evaluator should make an estimation of the contribution to general trends at programme area level (baseline trend), where feasible/statistically significant compared to other factors. Where this is not possible the evaluator should make a qualitative assessment in general terms.¹⁶

3.1.2 Second phase: Identifying the drivers of change

The second phase explores the contribution of specific measures and the ways in which the measure has contributed to (expected or unexpected) impacts in relation to specific context conditions or behaviour patterns of programme beneficiaries, non-beneficiaries and/or stakeholders. This phase provides the lessons of why and how the policy worked, in order to adapt or redesign rural development policies accordingly.

The CMEF requires measuring impacts at both the beneficiaries' (micro) and the sectorial and territorial (macro) level, in this very methodological sequence. Micro and macro level are linked by the intervention logic which provides a hypothetical trajectory from beneficiary over measure to objective and programme. In other words, the presumed chain of effects links the individual measures with the programme level. The following table provides a synopsis of how the seven common impact indicators are expected to be addressed by specific measures and axes according to the CMEF intervention logic.

¹⁶ Handbook on Common Monitoring and Evaluation Framework – Guidance Note A; DG Agriculture (2007)

Table I Rural development programme interventions and impact indicators

Objectives	Sub-objectives	Measure Code	Impact Indicators						
			Economic growth	Employment creation	Labour productivity	Reversing Biodiversity decline	Maintenance of high nature value farming and forestry areas	Improvement in water quality	Contribution to combating climate change
competitiveness of agricultural and forestry sector	knowledge and human potential	111	X		X				
		112			X				
		113			X				
		114			X				
		115			X				
	physical potential and innovation	121	X		X				
		122	X		X				
		123	X		X				
		124	X		X				
		125	X		X				
		126	X		X				
	quality of agricultural production and products	131	X		X				
		132	X		X				
		133	X		X				
	transition in new Member States	141	X		X				
		142	X		X				
environment and countryside	sustainable management of agricultural land and protect and improve natural resources	211				X	X		
		212				X	X		
		213				X	X		
		214				X	X	X	X
		215				X	X	X	X
		216					X		
		221					X		
		222					X		
		223					X		
		224					X		
		225					X		
		226					X		
		227					X	X	X
		228					X	X	X
quality of life	diversify the rural economy	311	X	X					
		312	X	X					
		313	X	X					
	quality of life in rural areas	321	X	X					
		322	X	X					
		323	X	X					
	territorial coherence and synergies	331							
Leader	Leader approach in mainstream rural development programming	411	X	X					
		412	X	X					
		413	X	X					
		421		X					
		431							

The socio-economic impact indicators are concentrated in the Axes 1, 3 and 4 with the impact on employment creation exclusively linked to Axes 3 and 4.

As can be seen as well, the four environmental impact indicators only apply to Axis 2 measures. However, as impacts have to be assessed at programme level, impacts originating from other measures will have to be taken into consideration as well. The same holds for economic impact indicators which are affected by Axis 2 measures.

In a similar way all baseline indicators are linked to the single objectives and sub-objectives¹⁷.

The measures can only take effect in mutual interaction with specific conditions of the social, economic and territorial context in which beneficiaries pursue their individual goals. It is one of the major aims of programme evaluations to highlight these conditionalities in respect to the national or regional context.

3.1.3 Third phase: Understanding change and concluding on future interventions

The interpretation of measured indicators and of qualitative (subjective and objective) information eventually allows for judging on the contribution of rural development measures to change and on the impact of the programme as a whole. In turn, these judgements provide the basis for proposing modifications of interventions, their respective budgetary endowment, or the policy approach as a whole. The third phase informs rural development policy as the ultimate goal of RDP evaluation.

The CMEF provides a set of common and horizontal evaluation questions which need to be answered, but additional programme specific evaluation questions are also important to widen and deepen the understanding of programme impacts in rural areas of the EU.

In the judgement phase, the evaluator answers all evaluation questions and draws conclusions from the analysis regarding the judgement criteria defined in the structuring phase. The conclusions and recommendations relate to the effects of single measures as well as the programme as a whole. The conclusions and recommendations should be strictly based on evidence of the quantitative and qualitative assessment. The limitations of the validity of the findings and the resulting judgement should be critically reflected.

The answer to each evaluation question must reflect the common and programme specific indicators. Where appropriate, other relevant information about the impacts of the actions in question needs to be taken into account. In all cases, the answers to evaluation questions must be accompanied by a critical discussion of the evidence of findings. Moreover, evaluation needs to consider the context within which measures are applied. If a certain measure or a part of the programme has not delivered the expected results and impacts, an analysis of the reasons for this unexpected effect is necessary.¹⁸

¹⁷ Handbook on Common Monitoring and Evaluation Framework – CMEF Guidance Note G; DG Agriculture (2007).

¹⁸ Handbook on Common Monitoring and Evaluation Framework – Guidance Note B; DG Agriculture (2007)

3.2 The Evaluation Architecture

3.2.1 Dealing with uncertainties

The whole process has to be understood as ideal-typical. However, while being applied, it involves many uncertainties:

- ▶ The data setup and information background may not be sufficient as to fulfil the requirements of quantitative indicator measurement (e.g. because of missing data, too small number of measurable units, the difficulty or impossibility to construct a valid control group of non-beneficiaries, etc.);
- ▶ Restrictions in carrying out qualitative surveys do not sufficiently allow for confirming the gathered information (e.g. the lack of willingness of non-beneficiaries to respond, the non-availability of key stakeholders for interviews, biased answering of questions touching aspects considered as private affairs, lack of resources, skills or sufficient time to go in depth of certain analyses, etc.);
- ▶ As a consequence, the chosen methodology will in many cases be a “second-best solution”, based on a trade-off between what should be done and what can be done;
- ▶ However evident evaluation results might appear to be, and how many times they might be confirmed by qualitative statements, the judgement is still a decision based on the evaluator's world view and the consequent filtering process. Any judgements require a radical reduction of complexity with all the risks to delete or overemphasize certain aspects of perceived reality.

3.2.2 Reducing complexity through a consistent approach

It will be impossible to get any information needed to be able to achieve fully confirmed evaluation results. They will be at best optimized approximations featuring a high degree of probability. The reasons are manifold but they can be summarized in the statement that the complexity of the system under observation (the impact of policy interventions on rural areas) is too complex to be grasped comprehensively.

In order to reduce the inherent complexity in a meaningful manner – which means without deleting or distorting essential aspects – the evaluators should aim at attenuating the complexity by interlinking all necessary steps and elements to each other within a consistent methodological framework, which we call evaluation architecture.

The evaluation architecture should encompass the three phases, connecting each step to (i) the overall purpose, (ii) the expected results and (iii) the quality requirements set by the European Commission and the programme level.

Within this evaluation architecture, the assessment of impacts

- ▶ follows the track from each measure to its potential contribution to (intended or unintended, wanted or unwanted) impacts (bottom up);

- ▶ helps to identify the individual, combined and synergetic contributions of different intervening factors (among those programme measures) to a selected set of impacts as addressed in the evaluation questions (top down).

Taking the inherent complexity into due account, the quality and usefulness of the evaluation will be better assured if the evaluation architecture comprises

- ▶ a mix of quantitative and qualitative approaches and methods within a coherent overall architecture;
- ▶ taking into consideration and cross-relating multiple perspectives (internal/external, different stakeholders) and world views (scientific/non-scientific);
- ▶ providing learning space and triggering learning processes among beneficiaries, intermediaries and programme authorities.

An illuminating example for consistent evaluation architecture is provided by IDEA Consult working for the Flemish Managing Authority for evaluating the impact of the Rural Development Programme 2000–2006. In order to assess the programme impact as comprehensively as possible, the evaluation was structured along 16 key themes, combining quantitative analysis with in-depth survey and expert workshops. This example is described on the following pages.

Impact assessment of rural development programmes ... the Flemish experience¹⁹

I. General approach

A. A vertical approach: effect themes as a starting point for impact assessment

The evaluation team opted for a vertical approach, in which the various measures were analysed for their (partial) contribution to the **overall programme effects** regarding a number of specific core themes/objectives such as income, employment, soil erosion, water contamination, etc.

The possible effects are classified into **three thematic components**. Within these, 16 relevant themes were distinguished for evaluation:

- agro(food) economy: income, employment, labour productivity and quality within agriculture, horticulture and the agro-food sector;
- the environment, nature and the landscape: soil, water, flora & fauna, landscape, ammoniac, greenhouse gases, water consumption and nuisance;
- rural quality of life: indirect employment, physical facilities, activities & services and perceived value of the physical environment.

B. A differentiated approach

The contribution made by various relevant measures to each of these aspects was evaluated on the basis of the **proportionality principle**: the main research efforts were devoted to those measures which were expected to have a high **impact**. This 'expected impact exercise' has been done on the basis of an interactive workshop, the available take-up information and existing causal-relationship-evidence (from *not necessarily RDP-related*) empirical research. Based on this exercise, those (combinations of) measures and impact themes were selected for which impact assessment has been carried out. This impact assessment exercise was **case-study-based** and at **micro/local level** – in order to detect net effects about RDP's

¹⁹ Katrien Van Dingenen, IDEA-Consult, Current Practice Example presented during European Evaluation Network workshop on assessment of socio-economic indicators, Rome, 28 October 2009

contribution to e.g. productivity.

For measures with a high potential (theoretical) impact but a low expected impact (e.g. due to the low number of project applications or for budget reasons), the main focus of attention was on implementation aspects on the part of the government and beneficiaries. This 'first level' analysis, which was mainly focussed at **outputs and results**, was done on the basis of the analytical framework of the European ITAES project – where meant/theoretical causality, targeting, participation, additionality and implementation issues are systematically mapped for each measure.

II. Theme-specific approach for axis 2 – labour productivity

The very basis of our analysis on the themes of productivity/income was the structuring and use of a diverse set of complementary data and information sources, being:

- An **expert-workshop** to identify and validate the most important (potential/expected/identified) causal relationships (~gross effect) between the RDP-measures and the themes concerned
 - Identification of theoretical/potential impact of specific (sub-)measures within the programme on productivity (all necessary circumstances such as sufficient budget & take-up, viability of farms,... fully present)
 - Estimation of the *level of* expected impact (5 category-level)
 - Validation of methodological assumptions and draft outcomes

This workshop was a combination of both plenary and thematic sessions (according to the axes of the RDP), involving both academic staff & sector experts and public administration personnel from different departments.

- A widespread **survey** among a representative sample of Flemish farmers, with questions on:
 - Which investments they made, or which agro-environmental schemes they adopted, and whether...
 - they did it with or without RDP-support
 - [if support was obtained,] whether (to what extent) they would have done it without RDP support
 - [if no support was asked/obtained,] why this was the case
 - Which effects these investments had on the labour conditions, production techniques, productivity etc in their farm (using 5 categories, ranging from negative influence to a major positive influence)
 - What triggered their decision to do this investment/to follow a training/to adopt a agri-environmental scheme/to diversify their activities (only one of them being RDP support) (~contextual analysis ⇒ net effect)

This survey enabled us to have semi-quantitative & structured information on issues, and at a level of detail that is not present in any database or institution. This source can be seen as a more subjective source of information – but with high added value as it gives very focused input for our evaluation – at micro/individual level.

- The accounting data in the **FADN** database for these (surveyed) farmers, with information on:
 - Employment & working hours per activity
 - Added value created
 - Output/turnover
 - Other farm-specific characteristics (type, ESU, geographical situation,...)

We worked with a sample for dairy farmers for more detailed analysis within a uniform group of farmers – which could be divided in 3 groups: (i) farmers having received RDP-support, (ii) farmers with investments but not supported by a public authority and (iii) farmers without any major investments during the examination period.

- The **accounting data** in an IDEA-owned database called Belfirst – for the agrifood industry, with similar scope of information as for the FADN database
- The **database of the Flemish Agricultural Investment Fund** – indicating in a high level of detail the type of investment for each farmer that made use of this Fund

– The **application dossiers** for the agrifood-measure (9.3.7 during RDP '00-'06)

The coupling of the accounting databases (FADN & Belfirst) allowed us to compare farms/companies with RDP-support and others – to detect correlations between e.g. investments and support (*what is the importance of investment support and the actual investments made?*); and between investments (with and without support) and evolutions in added value created per FTE etc (*what is the (micro level!) impact of these investments on the RDP-objectives?*).

These correlations at the level of farm investments could then be compared with the results of the survey. The survey also complements these data sources for the measures related to agri-environmental schemes and educational projects.

The result of this analysis was a **focused** but **cross-measure** and **micro level** assessment of potential and identified impacts of the Flemish RDP.

3.2.3 Constraints in utilization of the evaluation results

The evaluation report is just the beginning of an “added value chain” which should end up in improving policy orientation, structures, processes and rules of programming. However, the value of evaluation outcomes depends on a number of institutional challenges:

- ▶ Member States and the European Commission have to manage the filtering of very large amounts of information; at the same time the policy cycle revolves more quickly than the research cycle²⁰; institutional learning requires continuity of both policy contexts and organizational competences, and this is not taken for granted.
- ▶ The relative importance of certain evaluation aspects may change due to shifts in the policy debate (e.g. climate change was not considered as a relevant evaluation theme ten years ago).
- ▶ Financial or time-bound restrictions will enable or discourage managing authorities and evaluators to opt for more or less thorough evaluation designs.
- ▶ The evaluators with their specific competencies as well as the relationship between managing authorities and evaluators affect evaluation and assessment of impacts.
- ▶ Relating to the ex-post type of evaluations, the time frame poses problems. In most cases the planning for the next programme period is already in full run, which in turn limits the possibilities to capitalize on the results of the assessment of impacts of the previous period.

These are strong arguments to foresee strong links between monitoring, ongoing evaluation and impact evaluation over the whole programme implementation period, in respect to process steering, managing information flows and choosing the appropriate methodologies and instruments.

3.3 Assessing impacts of RDP – challenges and practical solutions

3.3.1 Methodological challenges

The main methodological challenges of a comprehensive impact evaluation are:

²⁰ Laschewski & Schmidt 2008

To answer the question: what would have happened to the respective programme area without a given programme?

The challenge is to provide evidence of a true cause-and-effect link between the observed indicators and the Rural Development (RD) programme. The difficulty to sort out the net effects by establishing a valid counterfactual is positively correlated with the relatedness of the observed phenomenon to space. This might be feasible for most Axis 1 measures, and more difficult for Axis 3 and Axis 4 measures. Finally, for many environmental impact fields (biodiversity, water quality, HNV, climate change) the counterfactual approach does not deliver sufficiently well. Due to the complexity and site specificity of potential environmental impacts of RD programmes, the identification of control groups and the establishment of a situation with and without the programme in place are very difficult. Moreover the lack of clear systemic borders of effects may lead to less reliable results in both the test and control groups.

To disentangle the effects of single measures or the programme as a whole from effects of other intervening factors:

The CMEF has established an unequivocal cause-effect chain for each measure. Following the intervention logic, environmental impacts would only be expected from Axis 2 measures – or socio-economic impacts only from Axis 1, 3 and 4 measures. However, the assessment of impacts is an integrative task. Cross-effects are most likely to occur, and the evaluation design has to take this into account.

Exogenous factors may have also influenced a given “impact indicator” calculated at the regional/macro level. The task to net out the programme-borne effects is specifically critical concerning environmental impacts, where there is still a lot of uncertainty about methods to determine the proper scale of appraisal (micro-macro). To a large extent the assessment of environmental impacts is therefore restricted to the measurement aspect – as in many cases the assessment of impacts is still in its infancy and a wide range of possible methods (such as in the field of socio-economic impacts) does not exist as yet.

To ensure the availability and validity of data and information required to construct a viable body of evidence:

With particular regard to the assessment of RD programme impacts on the environment the required data are often not available or have been collected for purposes other than establishing a cause-effect chain from a specific measure in the RD programme to the impact in an environmental compartment (soil, water, air). However the validity of these data mainly relies on long time series both for establishing a counterfactual situation and depicting any change due to interventions in a given area. This is due to the fact that environmental (unlike social) systems mostly follow much slower development paths, involving considerable time lags. Moreover, the question is to delimitate the systemic boundaries, i.e. the area within which observed effects (especially indirect ones) are deemed relevant.

To put the partial quantitative and qualitative results in a meaningful relation with the overall rural development programme and the overall policy context to be able to provide pertinent answers to the evaluation questions:

Again this challenge is more difficult to tackle for environmental than for socio-economic indicators. The focus of the assessment of environmental impacts as defined by the CMEF is challenging to depict the full range of rather vast, complex fields of extremely complex phenomena such as “climate change” or “biodiversity loss”. Moreover, the CMEF does not clearly distinguish *ecosystem functions* (photosynthesis, energy and matter recycling including respiration) from *ecosystem services* (relating to human and social interests such as productivity, storage or buffering capacity in respect to carbon, nutrients, pesticides, heavy metals etc., reducing ecosystems to “service providers” without intrinsic value). This conceptual gap presents further challenges in drawing conclusions.

3.3.2 Five key issues

The methodological challenges reveal five key issues which are dealt with in the following sub-chapters of section 3:

- ▶ The requirement to assess, wherever possible, the programme impacts against their counterfactual, i.e. calculating the changes that would have occurred without the specific programme intervention.
- ▶ The requirement to measure both the micro level and the macro level effects (using specific methodologies) and to meaningfully combine the results into one picture.
- ▶ The requirement to estimate the net effects of the programme, by netting out deadweight, substitution and multiplier effects.
- ▶ The requirement to construct a data and information base which allows for the unbiased computation of the effects as stipulated above.
- ▶ To bridge the gap between indicator measurement and a judgement of the functioning of the system as a whole.

3.3.3 The counterfactual assessment of impacts

3.3.3.1 Problems linked to the use of traditional (naïve) evaluation techniques

If there are no serious constraints hampering the construction of a valid control group, the calculation of the net effect of a given RD programme should not be carried out using traditional or “naïve” (statistically biased) evaluation techniques (e.g. simple before-after estimator). The main reasons are described here.

(a) Naïve before–after estimator (reflexive comparison)

The naïve *before-after* estimator uses pre-programme data collected for selected programme participants and compares them with the data collected for the same enterprises (programme participants) after implementation of the programme. The problem with this reflexive comparison is that the observed

changes in time $T_0 \Rightarrow T_1$ (before and after the programme) cannot be clearly attributed to the programme; they might have occurred anyway.

(b) Naïve comparison of programme participants with arbitrary chosen non-participants

Another naïve evaluation estimator uses all non-participants as a control group. While monitoring systems of RD programmes usually do not contain any information about non-participants, information about what would happen if programme beneficiaries had not participated in the programme is typically obtained from (rather ad-hoc) surveys carried out on selected outcome indicators (e.g. GVA, profits, employment, etc.) among those who did not participate in RD programme, irrespective of their level of similarity to programme participants, i.e. by disregarding systematic differences between these two groups. The approach relies on the assumption that in the absence of the programme the given outcome indicator (e.g. GVA per enterprise) for programme participants would be the same as for programme non-participants.

Clearly, this assumption would be justifiable if the average performance of programme participants were almost identical with average performance of programme non-participants. Yet, this is usually not the case. Since RD programmes (measures) often target specific groups (by setting eligibility criteria) programme participants either outperform or under-perform non-participants. If differences in performance of both groups prevail even without the programme, the **selection bias** in this estimator would be substantial.

(c) Naïve comparisons of programme participants with a population's average (Member State or region)

Another naïve estimator commonly applied in empirical evaluations of EU RD programmes uses a population's average (i.e. consisting of both programme participants and non-participants) as a control group. In this evaluation technique necessary data on the average outcome indicators (e.g. GVA per enterprise) in the group of programme participants and non-participants are usually obtained from various national statistics or surveys. The approach relies on the similar assumption as (b), i.e. in the absence of the RD programme the outcome indicator (e.g. GVA per enterprise) collected for programme participants would be the same as for an average of farms/companies population (programme participants and non-participants). This assumption however would only be justifiable if the average performance of programme participants (measured by any arbitrary outcome indicator, e.g. GVA, income, profit or employment) were identical with the averages computed for the mixed group of programme participants and non-participants.

(d) Use of performance standards

It can be also problematic to rely on average (regional or Member State) performance indicators/standards (baseline indicators) as proxies for a programme control group. Numerous studies showed that Member State's average common performance measures (e.g. average GVA in the sector) may not adequately represent a counterfactual situation (i.e. the situation of programme beneficiaries without the programme).

(e) Naïve standard DID (difference in difference) estimator

This estimator compares the before-and-after changes of selected result indicators (e.g. GVA per enterprise) for programme participants with the before-and-after changes of the same indicators collected for *arbitrary selected* non-participants, whereby the estimation of the effect of the RD programme is usually carried out on the basis of panel data involving a group of programme participants and an arbitrary selected group of non-participants. The DID estimator is already more advanced compared with techniques described above (a-d) as it assumes that the selection to becoming a beneficiary of a programme depends on both observables as well as on unobservables. Although in this method any common trends in the outcomes of programme participants and non-participants (fixed selection bias) get differenced out, the crucial assumption justifying this method is that selection bias remains time invariant (so called fixed-effect), and this is not often the case.

It can be easily shown that this estimator breaks if trends in the outcomes prove not to be time invariant, i.e. *in the absence of policy intervention* the differences between outcomes (e.g. GVA per enterprise) for programme participants and non-participants do not remain constant over time. Overall, the available evidence shows that standard DID estimators, though motivated by plausible stories about "fixed" differences in motivation, ability or performance, may be a poor choice in many evaluation contexts (Smith, 2000).

3.3.3.2 Recommendable evaluation design and methods for measuring the counterfactual

A viable solution: the quasi-experimental design

The counterfactual assessment of impacts requires quantitative analysis techniques for which, in principle, experimental, quasi-experimental and non-experimental evaluation designs are applicable.

However, experimental designs are based on randomly selected groups put in contrast to randomly selected comparison groups (the counterfactual); this constellation does not occur in the realm of rural development programmes, which makes it obsolete to go further into this matter.

It is therefore recommended to use the quasi-experimental design as the best possible solution and the non-experimental design as the second best one.

(a) The quasi-experimental design

Quasi-experimental (non-random) methods can be used to construct control groups when it is not possible to obtain participating ("treatment") and non-participating ("non-treatment") groups through experimental design. It consists of constructing a comparison group using matching comparisons.

Matching involves identifying non-programme participants comparable in essential characteristics to participants. Both participating and non-participating groups should be matched on the basis of either a few observed characteristics or a number of them that are known or believed to influence both participation and programme outcomes. Matched comparison groups can be selected before project implementation (through prospective studies) or afterwards (in retrospective studies).

Both groups, participants and non-participants, undergo before-and-after comparisons in order to measure change among those who have participated against those who have not participated. The design draws from available data sources and is therefore recommended for those measures which are targeting individual people, farms or enterprises which (i) can be aggregated up to a significant number and (ii) are not marked by exclusive features (space or community) which makes them somewhat "unique". Their main field of application will therefore be axis I measures.

For projects targeting beneficiaries with particular characteristics (e.g. the poorest ones, the farmers thought most likely succeed), it becomes difficult to find close matches for a comparison group. That is also the case with self-selection into the program – those who select to participate in the program may have different characteristics than those who are not willing to participate. The reliability of the results is often reduced, as the methodology may not completely solve the problem of selection bias. The selection bias can be statistically tested by applying balancing property tests (testing the similarity of covariates).

The matching methods can be statistically complex, thus requiring considerable expertise in the design of the evaluation and in analysis and interpretation of the results. As these methods usually involve complex statistical modelling, they often require abundant data, making the evaluation more expensive.

In case of full-coverage interventions such as nationwide policies and programs in which the entire population participates and if there is no scope for a control group, the before-after comparison (reflexive comparison) remains the only means to measure change.

(b) The non-experimental design

This evaluation design can be used when it is not possible to randomly select a control group (experimental design), nor to identify a suitable comparison group through matching methods (quasi-experimental design). In such situations, programme participants can be compared to non-participants using advanced statistical modelling to account for differences between the two groups. This evaluation also draws from existing data sources, but all in all it is more complex and statistically less robust than the quasi-experimental design. Full correction of the selection bias remains difficult. The enhanced data requirements of statistical modelling may turn out to be expensive.

Methodologies to construct control groups

For methodologies to construct control groups, firstly recommended is Propensity Score Matching (PSM), secondly Regression Discontinuity Design (RDD), thirdly Multivariate regression analysis; each is described in detail below.

The specific indicator sub-chapters (4.1-5.4) will refer to additional ones if appropriate.

(a) Propensity Score Matching (PSM)

PSM is a tool for identifying a suitable comparison group which can then be compared to the treatment group. Matching is done by using the **propensity score** which is defined as the predicted probability of participation given observed characteristics. This method allows one to find a comparison group from a sample of non-participants closest to the treatment group in terms of observable characteristics. The propensity score is estimated as a function of individual characteristics based on a statistical model (such as the logit or probit model). The project's impact is then the difference in outcomes between the treatment and comparison group.

In theory, PSM neither requires randomization nor pre-intervention data but in practice pre-intervention data is used to control for differences in farm/enterprise characteristics prior to implementation of a given RD programme (This is required if a combination of PSM and DID methods is applied). A second best is to use in the PSM post-intervention data only. Propensity score matching is very useful if there are many potential characteristics to match between a sample of programme participants and a sample of non-participants.

The sample size of beneficiaries should be large enough to ensure the high level of the statistical significance of the estimated parameters, but should normally exceed 50 observations.²¹ The sample size of non-beneficiaries should exceed the number of beneficiary observations by a factor 4 to 10. The more different beneficiaries and non-beneficiaries are, the larger should be the sample size of non-beneficiaries.

Apart from sample size, PSM requires a good quality of data: it implies the need to control all factors that influence the beneficiaries' response to a programme. In reality propensity scores will not be able to control all the differences between the two groups (treatment and comparison group).

(b) Regression Discontinuity Design (RDD)

RDD can be used when the programme foresees a distinctive eligibility threshold on the basis of one or more criteria. In simple terms RDD compares the outcomes of a group of measurement units (farms, people, firms) just above the cut-off point for eligibility with a group of units just below this point. As it is assumed that individuals around the cut-off point for eligibility are similar, the selection bias should be minimized. However, other complexities can also be influential.

RDD may be an adequate method where there are clear rules for project selection as being opposed to a programme targeted at a wide range of various beneficiaries.

²¹ The number of observations depends on interpretations of large size (higher than 1000) and small size (usually less than 50). PSM works better if a data consists of large size of observations. 50 is however rather arbitrary (it could be 45 or 40).

However, there are constraints. The RDD method assesses the marginal impact of the programme close to the cut-off point for eligibility, but very little can be said about units far away from it (in either direction). Furthermore, the threshold has to be really applied in practice; individual units should not be able to manipulate the parameter for becoming eligible.

(c) Multivariate regression analysis

Multivariate regression analysis is used to control for possible observable characteristics that distinguish participants and non-participants. If it is possible to control for all possible reasons why outcomes might differ, then this method is valid to estimate the programme effect. The differences in the mean outcomes of the two groups (participants and non-participants), related to the set of variables that cause participation and outcome, constitutes the programme or treatment effect.

The method takes into account all possible factors which may affect the outcomes, be they programme-borne or of different types and origins. However, to get valid evaluation results, (nearly) all possible influence factors have to be guessed beforehand.

Assessing the counterfactual: the DiD (DD) Method

The recognized method for measuring the counterfactual for assessing socio-economic impacts is the Difference-In-Difference/DiD (or: Double Difference/DD) Method.²²

This method can be used in the framework of an experimental, quasi-experimental and non-experimental design. DiD compares a **treatment and a comparison group before** (first difference) **and after** the intervention (second difference: difference in difference). The mean difference between the “after” and “before” values of the impact indicators for each of the treatment and comparison groups is calculated. The impact of the programme is the change in the value of the second difference compared to the first difference. This method can be in principle combined with propensity score method to adjust for pre-treatment differences that affect the parameter in question (e.g. economic growth, employed people).

DiD requires baseline and follow-up data from the same treatment and control group. The DiD delivers correct values if the selection bias is time-invariant, which means the method eliminates a selection bias that doesn't change over time. However, if a selection bias changes over time, the measured effects will be biased as well. So, the design is effective if the two groups have featured similar development paths in the past until the programme was introduced.

²² One distinguishes a simple DID (conventional approach) and a combination of ATT-and DID. DID design and DID method are used interchangeably.

There will be rarely those optimal conditions allowing for a full-fledged application of a DiD Method based on PPM. There are basically two limiting conditions:

- ▶ The first limiting condition is the scarcity of non-participants. If nearly all potential beneficiaries have indeed participated in the programme, it will be impossible to construct a computable control group:
In practice, a major problem in applying a with-and-without comparison is the identification of non-supported micro, small, medium or large scale enterprises. Often, nearly all investing firms in the region of interest got investment aid. And if they did not get any funding, the enterprises usually are basically different from those which received investment support²³.
- ▶ The second limiting condition is the uniqueness of participants. This specifically applies to spatially relevant measures aiming at impacts at community or local level, thus entailing a host of incalculable interrelationships.

3.3.3.3 Possibilities to deal with apparent non-availability of control groups

In some cases there are ways to circumvent this problem in relation to farm holdings:

- (1) First of all it is important to distinguish **current** programme beneficiaries from all other farms (i.e. including former programme beneficiaries). The rationality behind this split is rather simple: had the second group (i.e. other farms) in the examined period undertaken similar investments or had they economically developed in a similar manner as the first group (i.e. current programme beneficiaries) the direct effect of a given RD programme (e.g. 2007-2013) would be close to zero (i.e. due to high deadweight effects). In order to **enhance a better comparability** between both groups (current beneficiaries versus control groups, including former beneficiaries) one may also include into the list of covariates, i.e. the most important variables characterising the group of beneficiaries and non-beneficiaries prior to the given programme (variables to be matched on) a variable showing the “previously obtained level of support” (it is assumed that this variable affects both farm’s decision to participate and outcomes).
- (2) In rare situations where *in a given programme area* **all farms are beneficiaries** of a currently implemented (e.g. 2007-2013) RD programme (e.g. North Sweden) one can assess the effect of this programme in the framework of a **generalized propensity score matching** (GPS). The latter approach allows for **continuous** treatment (all farms can be supported from a given programme yet at **various** programme **intensity levels**).
- (3) Should above mentioned solutions appear as not feasible (e.g. due to insufficient institutional capacities or lack of knowledge of GPS methodology) one could change the level of analysis (e.g. from sub-regional NUTS4 to regional NUTS-3 or NUTS-2). The main rationality behind this is to increase the sample and thus the statistical probability of finding farms that were not supported by a given programme.

²³ Janetschek H. (2009), p.3 in reference to SME-related measures 311-313.

3.3.3.4 The counterfactual and assessment of impacts of project-type interventions

Generally, the assessment of impacts of project-type interventions (e.g. training, knowledge type of activities, etc.) within Axis 3 and 4 should be carried out at the level of community, region then up to programme area. Under the assumption that the intensity of a programme support depends on the characteristics of a given area/region where this measure is implemented, the assessment of the impact of a given RD measure can be carried out by applying similar methodologies as for Axis 1, except that instead of farms' characteristics the characteristic of a specific community, region, or programming area are taken as covariates. Furthermore, other synthetic outcome indicators, e.g. Rural Development Index can be applied to estimate the impact of above measures on the quality of life or rural population²⁴.

3.3.3.5 The counterfactual and situations of time lags and low uptake

The selection of an appropriate time period (after implementation of a given programme) may be crucial for estimating the programme results. Generally the period which is chosen should neither be too short (not yet unfolding outcomes) nor too long, as other confounding factors or policies specifically targeting either programme beneficiaries or programme non-beneficiaries may systematically affect estimated effects. As evaluation methodologies described above (PSM-DID or GPS) are quite flexible regarding the selection of an "end period", it is advisable to undertake re-estimations of outcomes by including the successive years, in order to verify the stability of the estimated impacts. Another possibility is to build 2 or 3 years averages as "end period". Unfortunately, there is usually a trade-off between advantages from re-estimations using successive years and disadvantages from worsening of the quality of the data base resulting from dropping units/observations from the balanced panel. The marginal effectiveness of such an exercise is to be estimated in each individual case by an evaluator.

Solutions to a low-uptake situation (low up-take at the beginning of the programme) can be similar as for slowly unfolding results. Solution is a re-estimation of results in the successive years (or building averages of years as end period). Another solution (a third best) would be to consider data collected from a year prior to a given unit's participation as a pre-intervention period (this may be different for individual units).

Regardless of these limiting conditions being relevant or not, the quantitative assessment should be complemented by qualitative methods for the following reasons:

- ▶ Qualitative surveys or case studies may complement the quantitative analysis where it does not deliver sufficiently significant answers on the question if the programme has been effective.
- ▶ Qualitative surveys are – apart from complex modelling (such as input-output models) – the only means to deliver answers to the question why and how a programme has worked to produce the measurable effects.

More about the crucial role of qualitative methods in sub-chapter 3.3.5.

²⁴ Michalek, 2009: as an example of a particular study, involving Poland and Slovakia, region- and time-specific factors (principal components, 991 variables in Poland and 314 variables in Slovakia) were thereafter used as covariates characterizing each individual region prior to implementation of a given RD programme.

3.3.4 Taking into account and cross-relating impacts at micro and macro level

It is clear that programme effects can not be expressed as the aggregated sum of individual effects, as this would mean to ignore the systemic character of economy and policy interventions. The question is how wide the scope of macro level impact assessment should be. The minimum requirement of a “hybrid” (both sectoral and territorial) programme such as the RDPs is to gauge the net effects at regional (specifically rural) and sectoral level.

Generally, the estimation of a programme's impact at a regional/programme level (e.g. via construction of the Net Additional Gross Value Added in purchasing power standard, or NAGVA-PPS, indicator) could be carried out using three alternative methodological evaluation techniques:

- (1) Statistical/econometric methods that control for the differences in:
 - (a) initial endowments and economic performance of programme beneficiaries (e.g. farms, food processing enterprises, specific rural communities, etc.) compared with equivalent non-beneficiaries
 - (b) conditions and policies undertaken in programme areas compared with non-programme areas (or with other areas characterized by a different intensity of a programme in question)
- (2) Regional input-output econometric models
- (3) Micro-macro models (including CGE framework)
- (4) System dynamics modelling (especially for assessing impacts in the environmental field)

3.3.4.1 Statistical/econometric methods controlling for differences at both micro and macro level

The first possibility to estimate an effect of the programme at the aggregated level is to use a micro approach by drawing on the principles of controlled experimentation (e.g. experimental or quasi-experimental approach) and to add up impacts measured at a sample level by using probability distributions of individual groups in the entire population. This can be done by measuring an individual response (individuals, households, farms, areas or regions) in controlled settings. Because the supported and comparison groups differ in observed and unobserved variables that affect economic outcomes, a simple comparison of outcomes between supported and non-supported units will not reflect the true effect of the programme.

To enable such comparisons various techniques can be applied to find **adequate controls** (e.g. matching; for details see propensity score techniques). An important step in this approach is to use respective outcome indicators (e.g. GVA per unit, GVA per labour employed, etc) and observable data on units in a given sample in order to derive a number of meaningful micro-based policy parameters, e.g. SATE (*sample average treatment effect*), SATT (*sample average treatment on treated*), STNT (*sample effect on non-treated*) and thereafter to estimate (by drawing on respective probability distributions) **aggregated** impacts for the population in large, e.g. PATE (*population average treatment effect*), PATT (*population average treatment effect on treated*), or ATNT (*population average treatment effect on non-treated*), see: Imbens and Wooldridge, 2007).

In many cases, PATE combined with information on programme costs (e.g. administrative costs and social costs) and general equilibrium effects (including substitution and replacement effects) can be helpful in answering the policy question regarding the net gain to the region, programme area or economy. Extension of this approach are techniques based on matched comparison of regional units (van de Walle, D., and D. Cratty. 2002; Lokshin and Yemtsov, 2005; Michalek, 2008). For example, in (Michalek, 2008) effects of RD programmes at regional levels were estimated using the Rural Development Index (RDI) – a proxy describing the overall quality of life in individual rural areas. The weights of economic, social and environmental domains entering the RDI index were derived empirically from the econometrically estimated intra- and inter-regional migration function after checking for alternative model specifications (i.e. panel estimate logistic regression nested error structure model, spatial effect models, etc).

The impact of RD measures implemented in specific rural regions was analysed by means of selected impact indicators in programme supported regions and control regions, prior to the programme and after it, by applying combination of the Propensity Score Matching (PSM) (e.g. Kernel matching) and difference-in-differences (DID) methods. Evaluation of programme results at regional levels was performed on the basis of the estimated Average Treatment Effects (ATE), Average Treatment on Treated (ATT) and Average Treatment on Untreated (ATU) effects using the RDI Index as the main impact indicator. Furthermore, sensitivity analysis (Rosenbaum bounds) was carried out in order to assess a possible influence of unobservables on obtained results. Given information on regional intensity to programme exposure (financial input flows) the overall impact of the support via a given RD programme (all measures) throughout the regions in a selected country was estimated by means of a dose-response function using the framework of a generalized propensity score matching (GPS), (Imbens, 2002; Lechner, 2002; Imai and van Dyk, 2002; Hirano and Imbens, 2004). An aggregate impact analysis proposed in (Michalek, 2008) allows also to test a number of common assumptions, i.e. impact of a given programme on the overall level of regional development or other indicators of regional performance, e.g. employment, labour productivity, environmental and social indicators, etc. In Michalek (2008, 2009) above methodologies were empirically applied for evaluation of the impact of SAPARD programme in Poland and Slovakia in years 2002-2005 at NUTS-4 level and effects of RDP programme (2000-2006) in Schleswig-Holstein (Germany). Results show a full applicability of the proposed approach to the measurement of the effects of rural development and structural programmes at respective regional levels (NUTS2-NUTS5).

Regional input-output econometric models

The second possibility to estimate an effect of the programme at aggregated level is to use standard regional input-output econometric models (e.g. REMI, IMPLAN, RIMS II or EMSI) applied in regional policy analysis to estimate direct, indirect and induced effects of a given RD programme. For example, REMI model, that has been in a continuous development since 1980s integrates input-output, CGE and economic geography methodologies. It consists of thousands of simultaneous equations and its structure consists of five major interrelated blocks: (1) Output, (2) Labor and Capital Demand, (3) Population and Labor Supply, (4) Wages, Prices, and Costs, and (5) Market Shares. The REMI model was applied in numerous studies of economic development in US and Europe, e.g. to evaluation of land use and growth controls, impact of investments in energy sectors, transportation, etc; evaluation of regional economic effects of investments in EU (Treyz F. and G, Treyz, 2002); and recently to ex-ante evaluation of RDP in Tuscany until 2020 (REMI-IRPET) (Felici, et. al, 2008).

Standard input-output modelling was also applied to the assessment of the Integrated Operative Programme (2000-2006) and in the ex-ante assessment of the Castilla y León (Spain) 2007-2013 Rural Development Programme. In order to apply I-O method, the increase in final regional sectorial demand originating from the intervention needs to be determined exogenously. This information is not directly observable and must be obtained from managers and coordinators of PDR programme or from direct beneficiaries of each measure. The main advantages of this method are: 1). The method allows calculation of the direct and indirect effects for the measures both in aggregate terms as well as for specific sectors., 2). The model may be applied for each Axes, for each year and for the entire Rural Development Programme over the whole period of execution (2007-2013). Therefore overall effects throughout the whole period as well as the differences between the respective Axes and years may be analysed., 3) The method provides prior estimations of the indicator based on forecasts in expenditure distribution, thereby offering advantages from the point of monitoring the Programme.

The main disadvantages of I-O methods are: 1) The method does not reflect possible changes in regional productive structure as it always uses the same Input-Output table. This proves to be even more of a drawback if the available input-output table offers data corresponding to a year lagged in time., 2) Quantitative and qualitative techniques need to be merged to determine expenditure distribution by areas of activity. The information required must take into account estimations of the various agents being in charge of executing the Rural Development Programme (Regional Government, Central Administration, Local Councils, Local Action Groups), meaning that it may prove quite difficult to obtain reliable estimates, and that the final results may be conditioned by the reliability of the responses provided., 3) The effects determined in this approach are only restricted to the individual region. Moreover, the method assumes that the effects are produced in a year or that the effect is accumulated. 4) The method does not take account of leverage effects and deadweight of the action, although the procedure may be improved by estimating side effects through surveys and case studies., 5) Due to the heterogeneity of the Programme itself, which covers a wide range of actions, the findings are restricted. In addition, certain measures are not aimed at boosting any particular sector of the economy but seek rather to impact the whole economy, as a result their quantification by sectors proves difficult.

Applicability of a range of models²⁵ the context of evaluation of EU policies raises several concerns. Firstly, it is not quite clear how a number of US economic parameters used in these models can be applied to EU reality, given different economic and social context in both economies (including problems with data classification and consistency) (comp. Wilson R. in: OECD, 2004); Secondly, modification of these models to reflect local circumstances is usually a considerable and highly time consuming effort that cannot be undertaken by a few external evaluators alone, but requires a great dose of cooperation with local authorities and local stakeholders; Thirdly, complexity of using models like REMI or LEFM undoubtedly requires a certain minimum level of expertise; Fourthly, problems with timeliness of the key data incl. input-output tables raises questions regarding forecasting validity.

²⁵ The IMPLAN model, recently extensively used for estimation of effects of economic US policies, is a computer software package that consists of procedures for estimating local input-output models and associated databases. Description of EMSI model is available in: Galloway, H. EMSI's Input-Output Model Multipliers: A Brief Overview and Comparison with Other Major Models, www.economicmodeling.com. Extensive comparison of multipliers used in REMI, IMPLANT and RIMS II models is available in: Rickman and Schwer, 1995.

3.3.4.2 Micro-macro models (including a Computable General Equilibrium/CGE framework)

The third possibility to estimate an effect of the programme at aggregated level is to integrate a micro-economic approach (e.g. micro economic individual behaviour or household models) into various local or regional models (e.g. Input-output, Social Accounting Matrix or CGE) and assess the impact of a programme on the base of these combined models (e.g. micro-simulation models with local/regional CGE, village CGE, etc.). Main advantage from the use of these models is a theoretical possibility to estimate both anticipated as well as non-anticipated effects; direct effects (at beneficiary level) and indirect effects (generated from supply of materials, goods and services attributable to other linked and not directly benefiting units and/or industries located in the same area as well as induced effects (i.e. multiplier effects) of a given programme generated through direct and indirect activities (including consumption, taxes, etc.) of a given policy (above models are subject to a consistency checks through micro-macro consistency equations).

3.3.4.3 Limits of micro-macro econometric methods and modelling²⁶

The main disadvantages of these models are:

- ▶ input-output models assume that technological/economic relationships are fixed over time and do not respond to price/cost changes;
- ▶ while input-output tables are normally available at relatively high aggregation levels their rescaling to a local level requires a usage of various (often non-transparent) procedures which can be divided in three main categories: “survey”, “non-survey” and “hybrid” approaches, e.g. location quotient approach (see: Del Corpo, et. Al, 2008);
- ▶ commonly applied CGE models usually do not show an enough detailed level of sector disaggregation (a major problem in evaluating RD policies) and are usually static (in contrast, multi-sector and regional dynamic CGE models are much more complex and time consuming in their building and therefore are very rarely applied to evaluation of policies at regional level);
- ▶ policy simulations based on CGE models are often carried out on the base of (constant) response coefficients (e.g. substitution elasticities) that are specific only for a given locality/region and therefore non-transformable to other regions;
- ▶ due to a high level of aggregation in “other sectors” many CGE models make an extensive use of exogenous variables in form of simple and highly predictable (e.g. linear) trends;
- ▶ in order to adequately represent specificity of a given region (and to maintain SAM data consistency) CGEs should to a greater extent apply coefficients (e.g. used to calculate various elasticities) that have been econometrically estimated using data from a given region and are not “borrowed” from other studies (often not comparable) and adjusted to reality through “necessary calibration”;
- ▶ empirical CGE modelling at regional level often is often impossible due to the lack of relevant statistical data at a local or regional level;

²⁶ Modelling is understood as integration of a micro-economic approach (e.g. micro economic individual behaviour or household models) into various local or regional models (e.g. Input-output, Social Accounting Matrix or CGE).

- ▶ the use of micro data sets to simulate the policy impact at macro level (common in micro-macro models) usually draw on the unverified assumption that a given sample data (e.g. individuals/units/companies at a local level) are statistically representative for an entire population (assuming a full integrability);
- ▶ in CGE models dynamics and stochasticity issues are usually non-sufficiently considered;
- ▶ in CGE modelling a heterogeneity of firm behaviour is largely ignored.

Despite these deficiencies, micro-macro models are increasingly applied to policy analysis and include the whole array of respective techniques, from the simpler macro models that use representative household groups to link macro economic policies with microeconomic data, to more complex top-down modelling frameworks that combine (top) macro models and (down) micro-simulation models (Bourguignon, et al. 2008).

3.3.4.4 System dynamic modelling

System dynamic modelling allows for a comprehensive systemic view in case of complex cause-effect chains. As the challenges in European regions are partly influencing each other and RD policies have to react upon these challenges, it is necessary to capture these cause-effect chains in a systemic way.

The complicated cause-effect relations are to be reduced into logic representation schemes and formulas which can be created with syntactically valid combinations of connectives, predicates, constants, variables and quantifiers. The cause-effect relations are then fit into the conception of the intervention logic of the different RD measures and in due course may be aggregated up to the Axis and Programme level.

System dynamics modelling²⁷ is a well-established method to depict and simulate complex system relations in a non-linear way, and allows for complex cause-effect relations and various feedback loops.

The modelling approach follows four steps:

- ▶ Conceptual model: the first phase of conceptualizing the model comprises the establishment of relations between all relevant model components (previous point) and the drawing of systemic borders. The elements of the model are to be selected carefully so that they show a direct relation to the system reality (in our case the causes and effects of EU directives on territorial impacts) and therefore allow for traceability for the user of the model, taking also into account the data availability. This implies that the number of status variables stays in balance with the available data (according to Jørgensen²⁸ the number of available data inputs shall amount to the second power (n^2) of the status variables n included in the model at minimum.
- ▶ Mathematical model: In this second phase of modelling, the relationships created within the conceptual phase are translated into formal model elements – thus formalizing the links between the status variables.
- ▶ Calibrated model: Now the adaption of the mathematical model to reality – resp. to the simulation environment has to be conducted. After having defined systemic relations necessary

²⁷ see Forrester J.W. (1971): Principles of systems; Wright-Allen Press; Cambridge, Mass. 2. preliminary ed., 5. print.

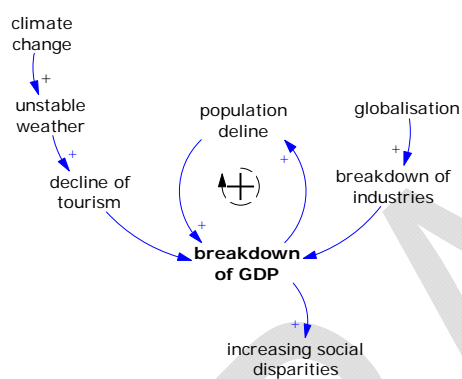
²⁸ Jørgensen B. (1993): The theory of linear models; Chapman & Hall; New York, NY [u.a.]

coefficients have to be included in order to – for example reflect effectively the empirical observations in the past.

- Dynamic model: After having completed these conceptual steps, the model is ready for dynamic modelling runs, with simulations of the changes within the model over time and with varying inputs (in our cases varying policy options).

The following two example pictures show such a modelling approach as conducted for the TERESA project. The project tried to identify typical interrelationships between farming activities, rural economy, rural society and the environment. Moreover it tried to identify and to assess different integration policies regarding their effectiveness in generating positive externalities for farming activities and rural development. The task within the assessment of impacts of RDPs is very similar, as it tries to depict regional challenges in all its cause-effect relations establishing links between the RDP measures and the territory.

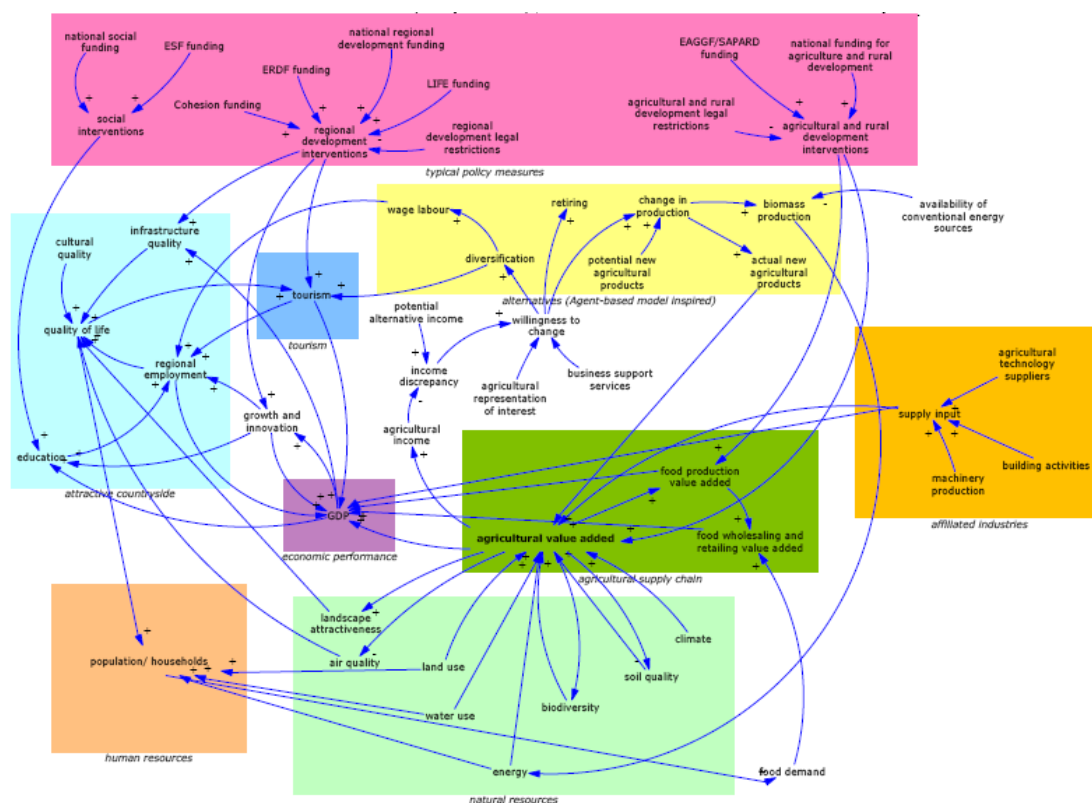
Figure 2 Causal loops connect different challenges



Source: ÖIR (2009)

The following graph shows the result of such a modelling approach establishing causal loops between policies, the actors and the different territorial compartments (environmental compartments, settlements, policy) of a region.

Figure 3 TERESA regional system dynamics model



Source: ÖIR (2009)²⁹

3.3.5 Netting out the programme effects by reducing deadweight, leverage, displacement, substitution and multiplier effects

3.3.5.1 Estimation of deadweight effects

According to the CMEF³⁰,

deadweight are changes observed in the situation of beneficiaries following the public intervention, or reported by direct addressees as a consequence of the public intervention, that would have occurred, even without the intervention.

Deadweight loss effects can be measured by using data for comparable farms (e.g. specialized in milk production) and applying a relevant result indicator (e.g. investment value per farm/enterprise and year; or value of inventories per farm/enterprise and year) for calculations of ATT (= average treatment on treated).

Estimation of deadweight loss at the level of direct programme beneficiaries can be carried out in the following way³¹:

²⁹ TERESA project (project conducted in the 6th RFP – <http://www.teresa-eu.info/>).

³⁰ Handbook of CMEF, Guidance Note N, p.3

³¹ by applying a method described in (Michalek, 2009)

- ▶ Identification of RD-programme supported units j carrying out investments under specific RD measure (e.g. Measure 1: Modernisation and Restructuring of Agricultural Enterprises)
- ▶ Identification in the control group (i.e. similar programme non-participants) a sub-vector of those farms m which undertook the same of type of investment as j (in period between $T=0$ and $T=1$)
- ▶ Calculation of ATT using data from both groups (i.e. j and m) and applying a selected result indicator (e.g. investment value per farms) before and after the programme
- ▶ Applying DID on the estimated ATT.

While it is expected that in case of a deadweight loss the calculated DID-ATT between above groups (j and m) will be close to zero, the estimated % of deadweight loss (between 0-100%) should be used to correct the estimates of programme direct effects (Step 1).

Example

Application of the above methodology to estimation of the deadweight loss effects in the RD Agrarinvestitionsprogramm in Schleswig-Holstein (Germany) (Measure: Investments in milk and beef sectors) on the basis of 1,333 bookkeeping farms (101 AFP participants and 1,232 non-participants) specialized in milk production (panel for years 2001-2007) indicates a considerable (93%) of deadweight losses (Michalek, 2009). Results are presented below.

Estimated deadweight loss effect of AFP programme on milk farms (Schleswig-Holstein, Germany)

Calculation basis	Value of inventories in EUR		
	2001	2007	DID (2001-2007)
Participants ($P=1$) (83)	80,058	153,545	73,487
Non-participants ($P=0$) (293)	57,379	108,539	51,160
Matched participants ($M=1$) (83)	80,058	153,545	73,487 (+92%)
Matched non-participants ($M=0$) (263)	70,181	130,733	60,552 (+86%)
Deadweight loss (M)			93% = (86/92)

3.3.5.2 Estimation of leverage effects

Leverage effect³² can be considered as an important micro-economic consequence of RD support. It signifies the extent of private spending which is induced by public funding (e.g. in form of RD programme).

The leverage effect can be calculated by taking the following steps³³:

- ▶ selection of individual units j supported by a RD programme
- ▶ identification of a comparison/control group k matching with units j (identical distribution of covariates) in the period $T=0$ (i.e. prior to j 's access to the programme)
- ▶ selection of relevant result indicators as proxies for private spending, e.g. money transfers from farm to farm households; level of private and farm consumption, etc.;

³² Propensity for public intervention to induce private spending among direct beneficiaries. In cases where public intervention subsidises private investments, leverage effects are proportional to the amount of private spending induced by the subsidy. Leverage effects must not be confused with additional effects (see net effect). (Source: CMEF guidelines, guidance note N, glossary)

³³ Michalek J. (2009)

- ▶ calculation of ATT for selected result indicators between both groups (i.e. j and m)
- ▶ Applying DID on the estimated ATT

It is expected that in case of a significant leverage effects the calculated DID-ATT will be positive and significant.

Example

Application of the above methodology to estimation of the leverage effects in the RD Agrarinvestitionsprogramm in Schleswig-Holstein (Germany) (Measure: Investments in milk and beef sectors) on the basis of 1,333 bookkeeping farms (101 AFP participants and 1,232 non-participants) specialized in milk production (panel for years 2001-2007) indicates significant leverage effects (Michalek, 2009), i.e. participation in AFP programme resulted in significant additional transfers of funds from farms to household (in average EUR 4,653 per farm see below; or EUR 3,178 per farm see below), thus induced private spending (see below)

Estimation of the leverage effects in AFP programme (Schleswig-Holstein)

Result indicator: Money transfer from farm to farm household for living

Calculation basis	Variable: Money transfer from farm to farm households for living		
	2001	2007	DID (2007-2001)
Unmatched P=1 (101)	30,072	43,810	13,738
Unmatched P=0 (1,232)	24,512	32,336	7,824
$\bar{0}$ (1,333)	24,933	33,206	8,273
Difference (1 minus $\bar{0}$)	5,560	11,473	5,913
Difference (1 - average $\bar{0}$)	5,139	10,604	5,465
Matched M= 1 (101)	30,072	43,810	13,738
Matched M= 0 (1,067)	27,647	36,732	9,085
ATT	2,424	7,077	4,653

Estimation of the leverage effects in AFP programme (Schleswig-Holstein)

Result indicator: Money transfer from farm for building of private assets

Calculation basis	Money transfers from farm for building of private assets		
	2001	2007	DID (2007-2001)
Unmatched P=1 (101)	18,447	48,302	29,855
Unmatched P=0 (1,232)	11,632	31,926	20,294
$\bar{0}$ (1,333)	12,148	33,167	21,019
Difference (1 minus $\bar{0}$)	6,814	16,376	9,562
Difference (1 minus average $\bar{0}$)	6,299	15,135	8,836
Matched M= 1 (101)	18,447	48,302	29,855
Matched M= 0 (1,067)	17,504	44,181	26,677
ATT	942	4,120	3,178

3.3.5.3 Estimation of general equilibrium (GE) effects (= substitution and displacement effects)

GE effects occur when a programme affects (positively or negatively) farms/enterprises other than direct programme participants. Important GE are substitution effect and displacement effect.

(a) Estimation of substitution effects

Substitution effects belong to macro-economic effects.

According to the CMEF³⁴, they are...

...effects obtained in favour of direct beneficiaries but at the expense of a person or organisation that does not qualify for the intervention.

For example, due to a given RD programme input/factor prices in affected region increase; or regional produce prices decrease compared with other regions. Substitution effect (in contrast to displacement effect) primarily occur in a *direct neighbourhood* of units j. It can be expected that this effect will influence all major programme result indicators, e.g. GVA per enterprise.

Substitution effects can be measured³⁵ by using similar techniques as in case of direct programme effects, yet comparing performance (change in GVA, employment or labour productivity) of programme non-beneficiaries in regions where intensity of a given programme exposure was high (high probability of positive/negative effects from a given programme) with performance of similar programme non-beneficiaries in other regions characterised by a low programme intensity. High difference in the estimated DiD-ATT indicates the occurrence of substitution effects.

Example

Application of the above methodology to estimation of the substitution effects in the RD Agrarinvestitionsprogramm in Schleswig-Holstein (Germany) (Measure: Investments in milk and beef sectors) on the basis of 1,333 bookkeeping farms (101 AFP participants and 1,232 non-participants) specialized in milk production (panel for years 2001-2007) indicates significant (negative) substitution effects (Michalek, 2009), i.e. implementation of AFP programme brought about some adverse effects (e.g. increase of a lease price of land) which lead to deterioration in situation of other farms (programme non-beneficiaries). Estimated negative substitution effects (in average EUR -3,546 per farm) are shown below.

³⁴ Handbook on CMEF, Guidance Note N, p.15

³⁵ applying a method described in Michalek, 2009.

Estimated negative substitution effects

Calculation basis	Profit per farm		
	2001	2007	DID (2007-2001)
Unmatched P=1 (526)	46,349	84,703	38,354
Unmatched P=0 (705)	40,531	83,034	42,503
\emptyset (1,231)	43,017	83,747	37,398
Difference (1 minus 0)	5,817	1,669	-4,148
Difference (1 minus average 0)	3,332	956	-2,376
Matched M=1 (517)	45,933	83,757	37,824
Matched M=0 (677)	48,559	89,930	41,371
ATT	-2,626	-6,172	-3,546

(b) Displacement effects

According to the CMEF³⁶, displacement effects are...

...effects obtained in an eligible area at the expense of another area. Displacement effects may be intended or unintended.

Displacement effects occur if farms *i* located in one geographical area (*ai*), which is not subject to RD support, becomes adversely affected by a programme support provided to farms *j* located in another geographical area (*aj*). For example, due to RD programme support to units *j* jobs are created in units *j* (located in programme assisted area *aj*) at the detriment of jobs *lost* in units *i* located outside of the area concerned (shift of jobs from *i* to *j* may lead to an increase of the generated GVA per enterprise in *j* and a decrease of GVA per enterprise in *i*).

Spatial displacement effects can be measured³⁷ comparing two relations: a) performance of programme supported units (*j*) with similar non-supported units (*m*) both located in regions characterised by a high programme intensity, and b) performance of programme supported units (*j*) located in regions characterized by high programme intensity with similar non-supported units (*k*) located in regions characterised by a low programme intensity before and after the RD programme. The lack of displacement effects would result in similar differences in DID-ATT between a) and b) (i.e. location of units would be considered as irrelevant).³⁸

³⁶ Handbook on CMEF, Guidance Note N, p. 3

³⁷ by applying methods described in Michalek, 2009.

³⁸ Generally speaking, and assuming no other general equilibrium effects (e.g. substitution effects), the bigger is the difference in DID-ATT between both groups (*j-k*) and (*j-m*) after the programme (the result of a shift of employment and a "shift" of GVA from units *k* to units *j* and *m*), the higher is the probability that the better performance of units *j* and *m* located in area *aj* occurred at detriment of units *k* located in non-supported areas *ai*.

3.3.5.4 Multiplier effects

Multiplier effects are defined in the CMEF³⁹ as...

secondary effects resulting from increased income and consumption generated by the public intervention. Multiplier effects are cumulative and take into account the fact that part of the income generated is spent again and generates other income, and so on in several successive cycles.

A correct estimation of programme multiplier effects is the most difficult issue. Generally, four different methods can be applied:

- (a) Micro-macro models (including CGE framework)
- (b) Regional input-output econometric models
- (c) Regional SAM models (Social Accounting Matrix)
- (d) Methods combining the Rural Development Index with the generalized propensity score matching⁴⁰

Modelling (a–c)

Among more sophisticated quantitative techniques applied to evaluations of the impact of new CAP policies the most often used were: I-O, SAM, General Equilibrium Modelling, e.g. CAPMAT,⁴¹ partial equilibrium and market projection models for the agricultural sector, e.g. CAPSIM,⁴² ESIM,⁴³ FAPRI,⁴⁴ and CAPRI.⁴⁵ While many of above models were constructed to provide information on *an aggregated effect* of certain policies and measures (e.g. production, consumption, trade, income, investment, etc.), their usefulness for ex-post evaluations of *RD programmes* is rather limited. *First*, most of these tools exhibit a rather high level of aggregation and do not allow, or only sporadically allow, for calculation of effects of a given policy on a more *heterogeneous group* of direct and indirect program beneficiaries.⁴⁶ *Second*, the above tools are not suitable to deal with a larger quantity of different economic actors. *Third*, they do not allow for the measurement of the impact of provision of non-marked goods. *Fourth*, the above type of modelling hardly allows for dynamic adjustments of regional structures.⁴⁷ *Fifth*, a large majority of the above models rely on rather unrealistic assumptions regarding individual formation of actors' expectations (Bourguignon and Pereira da Silva, 2003). *Sixth*, many important dynamic characteristic of economic actors' behaviour, e.g. consumption smoothing, labour migration (e.g. due to non-wage

³⁹ Handbook on CMEF, Guidance Note N, p.10.

⁴⁰ Michalek, 2009

⁴¹ See Keyzer et al., 2003

⁴² See EuroCARE, 2002a

⁴³ See EC, 2002d

⁴⁴ See FAPRI, 2002

⁴⁵ See EuroCARE, 2002b

⁴⁶ An enormous effort linked to building up a consistent data structure for a large number of actors prevents the user from applying those methodologies at regional or highly disaggregated levels, though in principle those tools could provide results at various disaggregation levels. Although the existing literature contains examples of general equilibrium models based on actual household surveys, the representation of unconstrained household behaviour and rather aggregated production side of those models are too simple to adequately represent required heterogeneity. Indeed, these models are usually adapted to a specific set of macroeconomic issues and not much to structural and sectoral problems that are typical of development (see Bourguignon and Pereira da Silva, 2003).

⁴⁷ Although some of these models may constitute a basis for a more comprehensive analysis of the impact of the RD programme at a regional level (e.g. CAPRI), their applicability for analysis of an impact of the programme on a larger number of different programme participants is strongly limited.

factors), or expansion of firms cannot be adequately modelled by applying the above techniques (any such modelling leads to an over-proportional and uncontrollable model complexity).

Generalized propensity score matching (d)

While the PSM methodology is helpful in analysis of a programme impact if some of analysed units/regions are exposed while some others were not exposed, a standard RD programme (i.e. in form of various measures) is normally implemented throughout the whole country, i.e. almost all regions have to be treated as beneficiaries. In this case, treatment (i.e. exposure to programme participation) is not more a binary- but instead a continuous variable. Because of this property, the classical setting using propensity score matching has to be extended. Propensity score techniques allowing for multi-valued and/or continuous treatment effects (i.e. generalized propensity score matching) proposed by (Imbens, 2002; Lechner, 2002; Imai and van Dyk, 2002; Hirano and Imbens, 2004) was applied to the evaluation of public contribution to enterprises in Italy (Bia and Mattei, 2007) and to the measurement of impact of RD programmes (SAPARD) in regions in Slovakia and Poland (Michalek, 2008).

The combination of PSM and DID methods (advanced DID application) is highly applicable in case the outcome data on programme participants and matched non-participants are available both “before” and “after” a programme in question has been implemented (Ravallion, 2004; Michalek 2008). A decisive advantage of the PSM-DID estimator is that by applying this methodology initial conditions of observable heterogeneity of both groups (programme participants and non-participants) that could influence subsequent changes over time are largely eliminated. In our view PSM-DID methodology is also highly applicable to assessment of programme impact at regional levels (under condition that appropriate controls (programme areas or regions) can be built. One of the difficulties commonly faced during formulation of a relevant base-line is a problem of a perfect comparability (ideally, in case of rural development programme, the same regions which participated in the programme should also be used for simulation of their performance without the programme). As this is not feasible, it is important to make comparisons in a manner which guarantees that basic characteristics of regions (expressed in terms of various regional indicators) used for such comparisons are as much as possible identical with those regions in which the RD programme was implemented (i.e. the statistical probability of receiving support from RD programmes should be the same for supported and non-supported regions in each comparison group).⁴⁸

3.3.6 Notes on data collection and processing

3.3.6.1 Quantitative and qualitative data: an overview

Indicators can be quantitative and objective, qualitative and objective, as well as qualitative and subjective. The following table provides an overview on the availability for different quantitative and qualitative indicators.

⁴⁸ Recently developed advanced evaluation methodologies (incl. PSM, and PSM-DID) were successfully applied in a number of studies that focussed on the measurement of effects of various structural, social and rural programmes in a number of countries, e.g. Dehejia and Wahba, 2002 (US); Newman et. al. 2002 (Bolivia); Venotoklis, 2004 (Finland); Jalan and Ravallion, 2001 (Argentina); Lechner, 2002 (Switzerland); Larson, 2000 (Sweden); Pradhan and Rawlings, 2002 (Nicaragua), studies focusing on evaluations of social funds projects and other programmes aimed at eliminating of poverty (Rawlings and Schady, 2002; Walle and Cratty, 2002; Bourguignon and Pereira da Silva, 2003; Ravallion, 2004). Yet, until recently their application to evaluation of EU RD programmes was only sporadic (Schmitt et al. 2004; Pufahl and Weiss, 2007; Michalek, 2008).

Table 2 Availability of data by source and geographical level⁴⁹

Type of indicator		Type of source	Data feasibility/availability			
			Local	Regional	National	European
Quantitative indicators	Objective/subjective	Compilation of physical and administrative data	+	+++	+++	+++
		Regular statistical analysis	+	++	+++	Eurostat
		Surveys and empirical research	++	++	++	++
Qualitative indicators	Objective (factual assessment scale)	Specific research (case studies, in-depth analysis)	+++	+++	+++	+++
	Subjective (personal assessment)	Opinion polls, interviews, focus groups, heuristics	+++	+++	+++	Euro-barometer

Notes: + difficult to obtain; ++ available; +++ readily available

Using quantitative indicators has some benefits like their measurability, the ability to aggregate data, to repeat measurements after a certain time span and to compare them to earlier ones. The downside of an imbalance on the use of quantitative methods is the normative power of what is measurable and what gets measured. The focus of the observer gets narrowed down towards the observable – and observed – variables.

While quantitative methods produce data that can be aggregated and analysed to describe and predict relationships between indicator values, qualitative research can help to probe and explain those relationships and to explain contextual differences in the quality of those relationships. Conversely, if qualitative research reveals interesting, surprising and sometimes counterintuitive relationships and patterns, quantitative research could be used to confirm or falsify the hypotheses inspired by those findings. The use of both describing and explaining provides the key to effective combination of methods and data. Qualitative and quantitative methods and data are often more powerful when combined, at different levels and in different sequences.

⁴⁹ Council of Europe 2005

3.3.6.2 Data and information sources for measuring the common impact indicators

The application of evaluation architectures involving the quantitative assessment of the counterfactual requires abundant data.

The FADN (Farm Accountancy Data Network)

FADN is the most relevant source of farm-related data. FADN and additional national farm accounting systems provide data on beneficiaries and non-beneficiaries for agricultural RD measures such as farm investment support (121), LFA schemes (211, 212) and agri-environment (AE) programmes (214, 215). The suitability of FADN data for RD evaluation could be enhanced if different types of RD support would be further disaggregated (e.g. differentiation of various agro-environmental measures or investment types).

Additional statistical data and surveys

Micro-statistics and periodic or evaluation-related surveys may complement these data, specifically for the non-farming sector, because outside the agricultural sector, the availability of secondary data becomes scarce.

Various Eurostat and national statistics on different economic sectors or households (e.g. labour force surveys⁵⁰) and enterprises are available, but availability varies strongly among Member States. In many cases, own surveys will be the only source of information for the evaluation of non-agricultural RD activities.

Purchasing Power Parities

The PPPs are compiled by Eurostat on an annual basis. Twice a year, in June and December, the most recent price survey results and national accounts data are incorporated into the calculation. The PPPs, PLIs and volume indices for GDP, its main sub-aggregates and a selected number of so-called analytical categories are published in Eurostat's dissemination database⁵¹.

Additional indicators for assessing environmental impacts

In the field of assessment of environmental impacts special emphasis will have to be put on the establishment of additional indicators (baseline and impact indicators). Programme specifics will call for the creation of additional indicators covering the territorial/environmental conditions of the programming area (e.g. specific species, soil conditions etc.) The availability of data (or rather the lack of it) will call for the use of the data which is readily available and collected under other circumstances (see e.g. the compulsory green-house gas inventory, which has to be provided by each MS – even on a regional scale – to be used for assessing the climate change impacts).

⁵⁰ Eurostat (2009): Labour force survey in the EU, candidate and EFTA countries. This publication describes the main characteristics of the national surveys. Eurostat also produces working papers on the topic: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-08-007/EN/KS-RA-08-007-EN.PDF

⁵¹ PPP data are available on the Eurostat website => Economy and Finance => Data => Prices. Source: Eurostat; Consumer Price Index (by regions: NUTS2 or NUTS3), or National Statistical Office, http://epp.eurostat.ec.europa.eu/portal/page/portal/purchasing_power_parities/data/main_tables

3.3.7 Bridging the gap between measuring impact indicators and providing answers on programme impacts

3.3.7.1 Notes on the complementarities between quantitative and qualitative approaches

There are two basic impact evaluation questions:

- ▶ To what extent did the policy work?
- ▶ Why and how did the policy work?

Impact evaluation of rural development programmes should give answers to both questions, as its rationale and purpose is to improve programme performance and inform and improve rural development policy.

Quantitative methods have comparative advantages to provide answers to the first question, and a mix of quantitative and qualitative methods may better be used to explore the second type of questions. A quantitative approach to assessment of programme effects that is based on analysis of observables should be combined with a qualitative analysis of un-observables that may have affected the obtained results. Good impact evaluation necessitates the combining of both qualitative and quantitative approaches to data collection and analysis, e.g. quantitative data collection (quantitative approach) with content analysis (qualitative approach), or inductive/naturalistic inquiry (qualitative approach) with quantitative measurement and statistical analysis (quantitative approach). Qualitative methods can also be used to better determine the design of quantitative surveys.

However, the distinction between quantitative and qualitative approaches is sometimes difficult to draw. Recent developments in theory-based modelling on one side and computational methods for content analysis on the other tend to blur the borderline between quantitative and qualitative methods.

Quantified research produces data in the form of numbers. Qualitative research tends to produce data that are narratives, images, or, at best quantified on the basis of factual assessment scales or personal assessments (rating scores). However, there is a difference in “objective” (*reality of the first order*) and “subjective” (*reality of the second order*)⁵² data, even if they are both expressed in (cardinal or ordinal) numbers. *Reality of the first order* is based on empirical evidence that can be repeatedly observed and confirmed. It provides the basis for scientific research in natural sciences. *Reality of the second order* emerges from subjective experience, values, perspectives and viewpoints. Reality of this kind is negotiated, the product of a *conversation*.⁵³ It is accessible to quantitative analysis, although only under certain conditions (e.g. large numbers of answerers in the case of opinion polls).

Qualitative approaches comprise interviews, focus groups, participant and direct observation (in the first case the researcher has to become a participant in the context being observed, and in the second case the researcher is just observing the objects without becoming one of them), and all kinds of case studies which may include hermeneutic analyses of behavioural patterns, interpretative analyses of narratives e.g. about critical incidents⁵⁴, and mindsets.

⁵² Watzlawick P. (1977): How real is real? Confusion, Disinformation, Communication. Vintage Publications, Millers Falls, MA.

⁵³ Gadamer, H.-G. (1979) Truth and Method, London: Sheed and Ward.

⁵⁴ Chapanis, A. (1959) Research techniques in human engineering, John Hopkins Press, Baltimore: Maryland.

Qualitative and quantitative methods are not only complementing each other, they need each other to unfold their potential value:

- ▶ Quantitative analysis sometimes requires qualitative methods (inquiries) to gather data or information. For instance, for computing gross added value in SMEs, it is necessary to ask the supported (and, for the counterfactual, non-supported) enterprises to provide the relevant numbers.
- ▶ Qualitative analysis helps to bridge the gap between indicator measurement and the interpretative answer to the common and horizontal evaluation questions (and possible programme specific evaluation questions), because they help to contextualize regulative ideas such as “quality of life” or “attractiveness of an area” (most relevant for axes 3 and 4). These terms only make sense if applied to a certain area and in the perception of those (e.g. inhabitants or tourists) who attribute emotional values to otherwise “neutral” features as are the untamed course of a creek, old fruits scattered over pastures and meadows, or the rhythm of seasonal feasts. The bundling of impacts into 16 relevant themes as was shown by the Flemish example (see section 3.2.2) provides a useful illustration for how the use of quantitative and qualitative, subjective and objective indicators can be intertwined into one consistent evaluation architecture⁵⁵.
- ▶ Qualitative analysis may also help to model possible relationships between the items addressed by indicators and by the evaluation questions; for instance, how the increase of labour productivity on a farm effects the quality of life of the people who work and live on this farm. Would this just mean to put people out of work? Or is that an opportunity to generate new non-farming employment based on the farm? Another example may be training measures: the path from participating in a vocational training course (output indicator) to an increase of GVA on the farm (result indicator) requires substantiation. Under which circumstances does it occur? What are the essential conditions in respect to the training method, in respect to consultancy or coaching after-training, in respect to the financial situation, the social and familiar background, the embedding in the community?
- ▶ For assessing environmental impacts qualitative information plays a pivotal role. Due to the specific reliance on qualitative and additional information for the assessment of environmental impacts there is the need for utmost transparency of the assumptions and of unequivocal communication flowing from the evaluators to the addressees of the evaluation.

In any event, qualitative methods provide valuable insights helping to understand the findings of quantitative research. They allow to model cause-effect chains or mutual causation cycles. If a program has proven to generate positive impacts, qualitative methods help to sort out what works and how it works, or to understand the circumstances in which unwanted side effects may occur.

Qualitative methods interlink different evaluation phases. *Prospective counterfactual impact evaluation* such as stipulated in the Report “*For a Reformed Cohesion Policy*”⁵⁶, starts from ex-ante evaluation, includes mid-term evaluation and finishes its cycle with ex-post evaluation. From the very beginning, the basic assumptions laid down in the intervention logic should be subject to systematic observation in order to

⁵⁵ We use the terms architecture and design here for similar purposes, although at different levels. The term “architecture” addresses the overall setup of RD evaluation, whereas “design” is used for the whole assessment trajectory for specific impacts or the assessment of impacts of specific measures.

⁵⁶ Barca F. (2009): *An Agenda for a Reformed Cohesion Policy*. EC DG Regio, Bruxelles.
http://ec.europa.eu/regional_policy/policy/future/barca_en.htm

be able to draw conclusions at a state when the implementation of the programme can still be substantially adjusted.

This implies that qualitative methods are a meaningful link between monitoring and evaluation. Systematic and multi-perspective observation architecture is the prerequisite of institutional learning at programme level.

3.3.7.2 Overview of qualitative evaluation methods

(a) Interviews

Interviews are the main methods for collecting data in qualitative research. The more structured the questions and the more limited the scope of possible answers, the more the results can be processed for quantitative analysis. Qualitative data ("second order reality", e.g. rating scores on the satisfaction of local partnerships with administrative procedures) can even be related to numbers stemming from quantitative ("first order reality") analysis (e.g. the average time from submission to approval of LEADER development strategies in a country)⁵⁷.

Semi- or unstructured ("deep") interviews are far more demanding in terms of the interviewers' questioning and interpretative skills. The resulting scripts can not be aggregated for statistical analysis, but they serve to generate, underpin or discard hypotheses about how the programme may have generated the effects assessed by quantitative evaluation.

Interviews can be used in various settings, such as for the hermeneutic analysis of behaviour and structural patterns⁵⁸, for the interpretative analysis of narratives (e.g. about critical incidents) and the exploration of stakeholders' mindsets. The computer-based repertory grid technique⁵⁹ combines the advantages of qualitative analysis (disclose implicit knowledge by exploring prevailing mind patterns) with those of quantitative analysis (computable and aggregable results, even over larger populations).

The results from interviews provide material to model cause-effect chains and mutual causation. A survey among farmers having undergone a training course may indicate how they utilised the acquired skills and translated them into added value on their farm. Deep interviews may reveal what factors favour or hinder the creation of enterprises in rural areas, what features are decisive for a positive valuation of a specific area as attractive.

On the other hand, interviews reach only a part of the programme-related population and depend on the willingness of addressees to respond. There is also the problem of the researcher's bias in the selection of the interviewees and by the way he/she poses the questions. This disadvantage does far less apply to the repertory grid techniques which in turn need to be computed by specialized software programmes⁶⁰.

(b) Focus groups

⁵⁷ An example for this approach is the *Ex-post evaluation of the Community Initiative LEADER II* (ÖIR Managementdienste 2003): http://ec.europa.eu/agriculture/eval/reports/leader2/index_en.htm

⁵⁸ Checkland, P.B. (2001) *Soft Systems Methodology*, in J. Rosenhead and J. Mingers (eds), *Rational Analysis for a Problematic World Revisited*. Chichester: Wiley

⁵⁹ Fransella, F./Bell, R./Bannister, D. (2004): *A Manual for Repertory Grid Technique*. 2. ed. Chichester (Wiley).

⁶⁰ Principal component analysis (PCA) with biplot diagrams

Focus groups consist of a small number of individuals brought together to discuss a topic of interest. Focus groups should be facilitated by an external moderator and comprise not less than 7 and not more than 15 people, as far as possible stemming from different sub-groups of stakeholders (e.g. managing authority, implementing body, beneficiaries, independent experts...). A focus group usually comes together several times; it can as well be installed as a core element for the continuance of an ongoing evaluation.

Thoroughly constellated and well facilitated focus groups may reflect the universe of programme stakeholders in a nutshell. Its added value is not only provided by the reflective capacity of the individual members, but by the group intelligence (*group intuition*)⁶¹ which brings forth new hypotheses and insights on critical factors and interrelationships, providing valuable lessons about patterns of success and failure. Focus groups are indispensable instruments for institutional learning (at the level of programme design and policy shaping). As such focus groups are an intervention in themselves, by strengthening the (self-)reflective capacity of a local partnership, specifically in the context of LEADER. As for impact assessment, focus groups produce a host of practical insights and help to sharpen the evaluation judgements; for example, the value added created by a direct marketing activity could be juxtaposed to the increase of work load on the female farming people, thus making a link between the gross value added and the quality of life indicator. Finally, the area-based perspective of focus groups implies that it could discover important realms which are not observed by the existing evaluation device. In this event it could propose and establish additional programme indicators.

As a matter of fact, focus groups involve only a small number of people and topics. It is therefore of utmost importance to pick the "right people", to find the right mix, and to avoid one-eyedness in the selection of participants.

(c) Case studies

A case study explores a specific situation or context in the form of a monography, but the possible kind, character and style of such as monography varies widely.. Case studies can be based on mainly quantitative analysis (such as for environmental impact assessment, but also for the comparative juxtaposition of costs and benefits of a programme-supported operation), on mixed forms (e.g. cost-effectiveness analysis, where the benefit side is at least partly expressed in qualitative terms) to genuine qualitative analysis (based on participative observation, deep interviews, interpretative analysis...).

Case studies serve to underpin or discard hypotheses, and due to their idiographic character they are rather useful to illustrate or to exemplify results which already seem to be sufficiently confirmed. With their help, intrinsic driving factors and cause-effect-relationships can be made explicit and thus become accessible for discussion. Case studies could, for example, help to assess the cross-effects of axis 2 measures on farm employment (axis 1) or the quality of life in rural areas (axis 3). They could also help to analyse which type of non-farming but farm-based employment (e.g. IT-based activities, tourism or social services) interrelates in which way with farming activities.

Due to costs, case studies also allow for studying only a small section of the programme-related population. There is also the problem of the researcher bias. As for socio-economic impacts, qualitative

⁶¹ Lukesch R., Payer H., Rabenau J. (2005): Wissen von innen – Fokusgruppen in der Begleitforschung zu Regionen Aktiv. In: Böcher M., Krott M., Tränkner S. (2005): Regional Governance und integrierte ländliche Entwicklung. Ergebnisse der Begleitforschung zum Demonstrationsvorhaben „Regionen Aktiv“. VS Verlag für Sozialwissenschaften, Wiesbaden.

research has its roots in social science and is more concerned with understanding why people behave as they do. Case studies involving environmental impacts pose the challenge of integrating natural sciences in an inter- or transdisciplinary setup.

(d) Process monitoring of impacts (short: Process monitoring)

Process monitoring of impacts⁶² is a method linking programme monitoring (which is mainly focused on input and output parameters) with the evaluation of the whole results chain up to the (intended or unintended) impacts. The results chain is constructed on the premises that the outputs generated through a development programme would produce the intended outcomes only if the beneficiaries make **use of the output** in the way assumed beforehand. The use of output is the variable linking outputs and outcomes. It is based on (usually mostly implicit) hypotheses on the expected behaviour of beneficiaries. For instance, if a farmer having attended a training course does not translate the acquired skills into alternative practice on the farm, the expected outcomes (e.g. increase of value added on farm) will not surface. If such gap is observed (which can be made within quite a short time) adequate measures can be taken (e.g. change the training curriculum, change the training institution, select the trainees in a different ways, provide coaching after-training, cancel the training offer at all and invest more in on-farm consultancy instead etc.).

The observatory function of systematic monitoring of the use of outputs may be accomplished by a focus group over the whole life cycle of the programme. Furthermore, it can be supported by interviews, case studies and other diagnostic instruments.

Impact monitoring may integrate various qualitative methods. It should be based on focus groups, but case studies and interviews may serve as additional sources of information. The continuous adjustment of the underlying assumptions allows for a flexible re-design of the programme approach and priority setting in full swing. This practice also increases the reflective capacity of the steering structure and contributes to embedding the lessons learnt into the institutions responsible for policy shaping and programme design. In other words, it is an investment in the social capital of an area, essential for the implementation of the LEADER approach.

Impact monitoring fully unfolds its potential as an element in a complex and consistent intervention architecture which has to be in place before the programme starts its operations. The requirement to translate the observed phenomena into neatly arranged and workable flow charts suggests the support from external experts.

⁶² Hummelbrunner R. (2005): Process Monitoring of Impacts. Towards a New Approach to Monitor the Implementation of Structural Funds Programmes. ÖAR Regionalberatung, Wien.

4. Guidance for assessing the socio-economic impact indicators and addressing the related evaluation questions

4.1 Impact Indicator 1: Economic Growth

4.1.1 The CMEF requirements

4.1.1.1 Definition of the impact indicator Economic Growth

According to the CMEF (Guidance note J – impact indicator fiches v1) the impact of a given EU RD programme on the economic growth (at a programme- and national level) is to be measured in terms of the indicator:

- ▶ the Net Additional Gross Value Added in purchasing power standard (PPS) – NAGVA-PPS.

The NAGVA-PPS indicator is defined as: “The change in the gross value added created directly in supported projects and indirectly in programme area that can be attributed to the intervention once double counting, deadweight, displacement and multiplier effects have been netted out”. Furthermore, NAGVA-PPS should be expressed in “purchasing power standards (PPS), by converting the net value added measured in euros or in national currency in PPS, using the conversion rates established by Eurostat”.

NAGVA-PPS indicator should be collected/estimated at the “national strategy level/programme level” in two consecutive steps:

- (a) Estimation “at level of direct and indirect beneficiaries by programme evaluator on the basis of output and result data, survey data and benchmark data and coefficients from similar projects and past evaluations (for calculation of double counting, deadweight, displacement and multiplier effects)”... after “cross-checking against counterfactual situation and contextual trends in programme area”.
- (b) Estimation of “contribution to general trend at programme area level (baseline trend), where feasible/statistically significant compared to other factors”.

The above indicator should be reported in: “ex-ante, mid-term and ex-post evaluations”.

NAGVA-PPS indicator should measure **impacts of a given RD programme on value added** generated in group of **direct programme beneficiaries** as well as other farms/companies **indirectly** affected by this programme.

The measurement of the overall impact of a given RD programme on **economic growth** using **NAGVA-PPS** indicator requires consideration of the following partial programme effects:

- ▶ **direct** programme effects occurring at the level of **direct programme beneficiaries** (direct effect of the RD programme on farm/company Gross Value Added (**GVA**) at a micro-level)

- ▶ **indirect** programme effects (e.g. deadweight loss, leverage effects, etc.) occurring at the level of **direct programme beneficiaries**
- ▶ **indirect** programme effects (**general equilibrium effects**) on **other individuals/farms/ companies** (programme non-beneficiaries) **affected** by a given RD programme (e.g. substitution effects, displacement effects, multiplier effects)

The major milestones in construction of the NAGVA-PPS indicator are:

- (1) Calculation of **changes in gross value added (GVA)** generated by **direct and indirect** programme beneficiaries in case of:
 - a) **implementation**, and
 - b) **not implementation** of a given RD programme
- (2) Computation of **net programme effects** (i.e. net additional gross value added) via **aggregation** (adding up) of differences between 1a) and 1b) for **all units** directly and indirectly affected by a given RD programme.
- (3) Expressing final results in purchasing power standards (PPS)

4.1.1.2 Related Common Evaluation Questions

NAGVA-PPS indicator should be used to answering various horizontal and measure specific Common Evaluation Questions (CEQs):

- (1) Horizontal CEQ (all Measures):
 - a) To what extent has the programme contributed to the realisation of Community priorities in relation to the renewed Lisbon strategy for growth and jobs with respect to:
 - Ameliorating the conditions for growth?
 - b) To what extent has the programme contributed to achieving economic and social cohesion policy objectives with respect to:
 - Reducing the disparities among EU citizens?
 - Reducing territorial imbalances?
 - c) To what extent has the programme design been successful in avoiding deadweight and/or displacement?
 - d) To what extent has the programme design been successful in encouraging multiplier effects?
- (2) Measure Specific CEQ
 - (a) Axis I

To what extent has the aid contributed to improving the competitiveness of the agricultural sector?

 - Applicable to Measures: 112, 121, 122, 123, 124, 125, 131, 132, 133, 141, and 142.
 - (Input to the calculation of result indicator for Measures: 113, 114, 115, and 126).
 - (b) Axis 3

To what extent have supported investments contributed to improving the diversification and development of the rural economy?

 - Applicable to Measures: 311, 312, and 313.

To what extent have supported investments contributed to improving the quality of life in rural areas?

- Applicable to Measures: 311, 312, 313, 321, 322, and 323.

(c) Axis 4 (Leader)

To what extent has the LEADER approach contributed to mobilising the endogenous development potential of rural areas?

- Applicable to Measure: 4

To what extent has the LEADER approach contributed to the priorities of axis 1, 2 and 3?

- Applicable to Measure: 4

4.1.2 Key challenges with regard to measurement and interpretation

Construction of the **NAGVA-PPS** indicator (expressed in **PPS** standards) requires carrying out the following steps:

- (1) Collection/Calculation of gross value added generated by RD **programme beneficiaries** at the micro-level (farm or food processors) in a selected programme area **prior** to the programme and **after** programme implementation.
- (2) Collection/Calculation of gross value added coefficients generated by **similar enterprises** (e.g. farms, food processors, etc.) which **did not participate** in a given RD programme (e.g. through matching) in a selected programme area.
- (3) Calculation of the **% change in gross value added** created at the group of beneficiaries **caused by the RD programme** (by deriving appropriate counterfactuals and computing Average Treatment on Treated Effects (ATT) using a combination of DID and ATT methods.
- (4) Explicit selection of **other** groups of enterprises considered to be **indirectly** affected by the RD programme in a selected programme area (e.g. other farms/enterprises, food processors, local input providers, etc.).
- (5) Calculation of the **% change in value added** in the above group (indirect programme affected: positively and negatively) and **caused by the programme** in a selected programme area.
- (6) **Aggregation** of the changes in value added in both groups (direct and indirect programme beneficiaries) in a selected programme area.
- (7) Calculation of RD programme **general equilibrium effects** (substitution, displacement, multiplier, etc.) in a selected programme area.
- (8) Calculation of the **net** additional gross value added in a given programme area by subtracting (7) from (6)
- (9) Calculation of (8) in all respective regions (programme areas).
- (10) Expressing (9) in purchasing power standards

In view of above, **the following key issues** are linked to the measurement of the impact of a RD programme on the economic growth using the **NAGVA-PPS indicator**:

4.1.2.1 Determining true causation

Determining true causation⁶³ means to verify that an observed (at micro- or regional level) change in **GVA** (per enterprise) that might be theoretically (!) associated with a given RD programme (RD measure) can indeed be attributed (as a whole or partly) to (or is caused by) this policy intervention. In order to verify above supposition, *effects of others intervening factors, (i.e. exogenously determined) which may also influence an observable outcome measure (GVA) have to be separated ("netting out") from effects of a given policy intervention.* Separation of programme effects from other factors requires a construction of an appropriate **counterfactual base-line (the key issue in evaluation of programme impacts!)**. In our case, the latter is expected to provide an answer to the question: What would be the **GVA** per enterprise in programme supported enterprises *without* a RD programme? Clearly, a counterfactual performance of supported enterprises **cannot be directly observed**. For the same reason, in non-experimental studies programme's causal effects (e.g. effect on **GVA**) should be assessed by making comparisons between supported enterprises with **appropriately constructed control groups** which did not benefit from the programme (see:2).

4.1.2.2 Elimination of a selection bias at micro- and macro-level

The selection bias in evaluating the impact of RD programme occurs if the mean outcome (e.g. **GVA** per enterprise) of those units which participated in RD programme *differs* from the mean outcome of the control group (non-supported units) *even in the absence of support*. An important problem which usually arises while simply comparing average data for programme participants and non-participants is that many RD programmes/measures are *not assigned randomly* but: i) are designed to target **specific beneficiaries** with a certain performance characteristic (e.g. under performed producers/enterprises/areas, etc.), or ii) include various eligibility conditions which, in practice, can only be fulfilled by certain types of economic units, e.g. the best enterprises. In both cases, a supported group may easily **outperform/under-perform** specific control groups or national averages, making simple comparisons of both groups' performance (e.g. **GVA** per enterprise, **GVA** per employed or **GVA** per ha) statistically biased thus unacceptable. Ideally, control groups (enterprises/producers) should differ in their basic characteristics from the supported group only in so far as they do not receive any intervention. To be meaningful, a control group should therefore consist of only those enterprises which *matched* in their observable characteristics with supported enterprises (*prior* to the programme).

4.1.2.3 Aggregation of various direct and indirect programme effects

Various **direct** and **indirect programme effects** (measured in terms of **GVA per enterprise**) estimated at the level of direct and indirect programme beneficiaries (e.g. in other farms/enterprises affected by the programme) have to be aggregated at a given programming area or country level.

Cumulative net change in **GVA** can be obtained by applying three alternative methods:

- ▶ **Method 1:** by adding up all separately computed **net** programme effects for **all relevant groups of farms/enterprises affected by the programme** (direct beneficiaries, indirect beneficiaries, other affected groups/sectors), or

⁶³ Causation cannot be proved through a simple correlation analysis.

- ▶ **Method 2:** by applying advanced evaluation techniques (e.g. generalized propensity score matching) enabling a direct estimation of **net** programme impacts on an **aggregated value added generated in specific sectors** (e.g. agricultural sector, food processing, local tourism, etc.) at regional- or macro level. By using this method, however some of indirect programme effects (i.e. deadweight losses and leverage effects occurring at the level of direct programme beneficiaries) would not be accounted for, or
- ▶ **Method 3:** by combining Method 2 with Method 1 in the sense that aggregated effects computed at a regional level (Method 2) are corrected by taking into account some programme indirect effects calculated at a direct beneficiary level (deadweight loss effects and leverage effects using Method 1)

4.1.3 Recommendable methods of measurement

Taking into consideration theoretical requirements for constructing the **NAGVA-PPS** indicator, the practical approach may involve the following stages:

Stage 1: Estimation of direct programme effects occurring at the level of direct programme beneficiaries (direct effect of the programme on GVA at a micro-level)

Direct programme effects at the level of direct programme beneficiaries can be estimated by using **GVA** per enterprise as the programme result indicator at a micro level (e.g. in case of agricultural farms, data about GVA can be obtained from FADN).

The following steps may be carried out⁶⁴:

- (1) Using information about general and measure specific conditions for programme participation select from available data base (e.g. FADN data) all those farms/enterprises that were programme eligible (for a given measure) prior to the beginning of the RD programme (measure specific selection)
- (2) Divide above group in programme beneficiaries vs. non-programme beneficiaries
- (3) Given information on **GVA** per enterprise and other important farm characteristics (e.g. land area, employment, value of assets, etc) **before** the programme ($T=0$), select from both groups **comparable** farms/enterprises (e.g. apply a matching method).
- (4) Check statistically the “similarity” of both groups **prior to** their participation in the programme (e.g. by performing balancing property tests on the most important **farm characteristics**)
- (5) Calculate specific policy indicators, e.g. Average Treatment Effects on Treated (ATT) to be estimated **before** the programme, using **GVA per enterprise ($T=0$)** as the result indicator
- (6) Collect data on **GVA** per enterprise for both (matched!!) groups of farms/enterprises (beneficiaries vs. non-beneficiaries) **after** implementation of the programme ($T=1$)

⁶⁴ Similar approach was applied, among others, to assessment of the impact of the SAPARD programme (2002-2004) on agricultural companies in Slovakia and the impact of the Agrarinvestitionsförörderungsprogram (AFP) on milk producers in Schleswig-Holstein (2001-2006). See: Michalek (2007, 2008, 2009).

- (7) Perform calculation of specific policy indicators, e.g. Average Treatment Effects on Treated (ATT) **after** the programme, using **GVA** per enterprise (**T=1**) as the result indicator
- (8) Apply **conditional DID method** (combination of ATT and standard DID) to calculate the **first component**, i.e. the net effect of the RD programme on **GVA** generated by **programme beneficiaries** (at micro-level)
- (9) Perform a sensitivity analysis of obtained results

Two empirical examples

Example 1

Above methodology was applied in (Michalek, 2009) in order to assess the impact of the SAPARD programme in Slovakia (measure 1: Investment in agricultural enterprises) on programme beneficiaries. The SAPARD support under Measure 1 was primarily targeting the following agricultural sectors: a) beef sector, b) pig sector, c) sheep sector, d) poultry sector, e) fruits and vegetables sector. Programme support under Measure 1 had a form of a capital grant covering up to 50% of costs to investments in the above sectors. The major beneficiaries of programme support (received approximately 67% of funds available under this measure) were large agricultural companies located in relatively well developed regions of West Slovakia (Nitra, Trnava and Bratislava). Assessment of the impact of the programme on agricultural companies was carried out on the basis of Slovak FADN data base in years 2002-2005.

The following steps were carried out:

- (a) SAPARD beneficiaries were identified and selected from the existing FADN databases to the panel. Data for each SAPARD beneficiary was collected in years 2002-2003 (i.e. prior to their participation in SAPARD) and 2005 (after implementation of SAPARD).
- (b) SAPARD general and specific eligibility criteria (e.g. pre-defined farm performance coefficients and farm profitability ratios; various minimum/maximum production-, age-, etc. thresholds; etc.) that were valid in individual years were translated into respective quantitative coefficients and applied to all non-SAPARD units included in FADN databases
- (c) Units which satisfied the above criteria in years 2002-2005 and did **not** receive a support from SAPARD programme and were selected to the panel as *eligible* non-participants
- (d) Respective balanced panels (i.e. embracing SAPARD beneficiaries and all non-SAPARD units meeting SAPARD eligibility criteria in specific years) were constructed for the years 2002-2005, i.e. observations on the same units in period 2002-2005.

On the basis of available Slovak FADN database, 232 agricultural companies were selected for the further analysis (balanced panel data) which was performed for the years 2003 (before SAPARD) and 2005 (after SAPARD)⁶⁵. Of selected 232 agricultural enterprises there were 51 agricultural farms SAPARD participants and 181 SAPARD non-participants (yet, SAPARD eligible).

The preliminary analysis showed that agricultural companies which received support from the SAPARD programme **differed** significantly in a number of important characteristics from eligible programme non-participants, i.e. SAPARD beneficiaries were in general much larger (ha), they employed more people and were more profitable (i.e. less unprofitable) compared with those agricultural companies which were non-supported from SAPARD. Given above, the group of SAPARD participants could not be directly compared with non-participants (selection bias!).

In order to ensure comparability, the propensity score matching (PSM) method was applied using number of individual characteristics of agricultural companies as covariates (methodology applied to selection of variables into the estimated logit function is described in this study).

Imposition of common support region and selection of appropriate matching algorithm resulted in dropping from a further analysis those companies which were non-comparable.

The applied balancing property tests (t-test) showed that the selected matching procedure (i.e. kernel epanechnikov bandwidth 0.06) considerably improved comparability of both groups of agricultural companies, making a counterfactual analysis more realistic. Indeed, previously existed significant differences in the most important farm characteristics between the group of agricultural companies supported from the SAPARD programme ($D=1$) and non-supported farms ($D=0$) dropped after matching (differences became no more statistically significant). This

⁶⁵ All selected beneficiaries received support from SAPARD in year 2004. Unfortunately, inclusion of the following years (2006 and 2007) was not possible due to dropping of many former agricultural companies from the data panel.

applies to all important variables determining both programme participation and outcomes, e.g. profit per company (prior to SAPARD programme), liabilities, value of current assets, etc.

The assessment of the micro-economic effects of a given RD programme on programme beneficiaries was carried out in both (now comparable!) groups of farms using seven results indicators available from a standard FADN system:

- Profit per company
- Profit per ha
- Profit per person employed
- Gross value added per company
- Employment per company
- Labour productivity (Gross value added per employed)
- Land productivity (Gross value added per ha)

The obtained results confirmed a high applicability of the above method to estimation of programme effects on beneficiaries of SAPARD programme. The results showed also that traditional estimates of programme effects can be highly misleading, whereas application of advanced evaluation methodologies can lead to quite different yet, much more reliable results.

This issue can be very well illustrated on the basis of the table below.

Estimation of the effect of SAPARD on agricultural companies (comparison of various methods)

Observations	GVA/company In SKK 1,000		
	Before SAPARD programme (T ₀)	After SAPARD programme (T ₁)	DID
SAPARD participants (P=1)	17,727	18,478	751
SAPARD non-participants (P=0)	9,950	9,680	-270
Average Ø	11,660	11,614	-46
Difference (1-0)	7,777	8,798	1,021
Difference (1- Ø)	6,067	6,864	797
Matched SAPARD participants (M=1)	11,082	9,610	-1,472
Matched SAPARD non-participants (M=0)	9,367	9,701	334
Average Treatment Effect on Treated (ATT)	1,715	-90	-1,805

Analysis of figures in the table above shows that:

- If a naïve before-after estimator (i.e. statistically biased) was applied the effect of the programme would be assessed as very positive (average change in GVA per company = +751 thousand SKK)
- If SAPARD participants were compared with **all** SAPARD **non-participants** (before and after) and DID estimator was applied the effect of the programme would also be assessed as very positive (average change in GVA per company = +1,021 thousand SKK). Yet, this estimator is statistically biased.
- If effects observed for SAPARD participants were compared with a **country's average** (e.g. **performance standards**) calculated for all farms, i.e. SAPARD participants and non-participants (before and after) and DID estimator was applied the effect of the programme would be assessed as positive (average change in GVA per company = +797 thousand SKK). Yet, similar as in (1) and (2) this estimator is statistically biased.
- Yet, above **conclusions** have to be **revised** in case the programme effects are measured using **statistically similar groups** (participants vs. non-participants). In this case the estimated programme effect (DID in ATT) was found to be **negative** (average change in GVA per company = -1,805 thousand SKK). The reason is a much higher growth in GVA per company in the matched group of SAPARD **non-participants** (average change in GVA = +334 thousand SKK) compared with the **matched** SAPARD **participants** (average change in GVA = -1,472 thousand SKK).

Example 2

Application of the PSM method (DID-ATT) to the evaluation of the RD Agrarinvestitionsprogramm (AFP) in Schleswig-Holstein (Germany) (Measure: Investments in milk and beef sectors) on the basis of 1,333 bookkeeping farms (101 AFP participants and 1,232 non-participants) specialized in milk production (panel for years 2001-2007) using profits as outcome indicator shows **positive effect** of the AFP programme on programme beneficiaries (average change of profit + EUR 10,661), yet much smaller effects compared with traditional methods (see:

(traditional methods vs. PSM methodology) are in the table.

Estimated effect of AFP programme on milk farms (Schleswig-Holstein, Germany)

Calculation basis	Profit per farm in EUR		
	2001 Prior to participation	2007 After implementation	DID (2007-2001)
Unmatched AFP participants P=1 (101)	54,629	116,777	62,148
Unmatched Non-participants P=0 (1,232)	43,007	83,718	40,711
Ø (1,333)	43,888	86,222	42,334
Difference (1-0)	11,621	33,059	21,438
Difference (1-Ø)	10,741	30,555	19,814
Matched AFP participants M=1 (101)	54,629	116,777	62,148
Matched Non-participants M=0 (1,067)	55,266	106,752	51,486

Stage 2: Estimation of indirect effects of a RD programme at the level of direct programme beneficiaries.

Methodologies recommended for estimating deadweight, leverage, substitution and displacement, as well as multiplier effects are described and illustrated by examples in sub-chapter 3.3.3.

4.1.4 Data requirements and collection

The necessary “data ingredients” for calculation of NAGVA-PPS indicator using above methods (**excluding multiplier effects**) are:

- ▶ Gross value added and data on other important farm/enterprise characteristics calculated prior and after implementation of a given programme for programme **beneficiaries** (panel data); Source: FADN, national farm accountancy network, micro-statistics, surveys.
- ▶ Gross value added and data on other important farm/enterprise characteristics calculated prior and after implementation of a given programme for **similar** programme **non-beneficiaries** (panel data); Source: FADN, national farm accountancy network, micro-statistics, surveys.
- ▶ Gross value added calculated prior and after implementation of a given programme for **selected programme non-beneficiaries indirectly affected by the programme** (panel data). Source: FADN, national farm accountancy network, micro-statistics, surveys.
- ▶ Gross value added calculated at sector level (Sections: A; A+B; C-E; F; G-I). Source: EU National accounts: Agriculture, other sectors.
- ▶ **Purchasing Power Parities** as compiled by Eurostat on an annual basis (see sub-chapter 3.3.4).

4.1.5 Interpretation and judgement issues

Once appropriately computed, the NAGVA-PPS indicator can be interpreted as the net effect of RD programme impact on the economic growth (measured in net value added PPS) in a given region, thus enabling respective policy conclusions.

The “explanation gaps” between indicator measurement and assessment of impact will depend on the number of factors. The most important are:

- ▶ Quality of available data
- ▶ Knowledge of a suitable methodology
- ▶ Extent and depth of analysis (e.g. pre-selection of sectors/farms/enterprises considered as **indirectly** affected by a given RD programme).

Despite many advantages the NAGVA-PPS indicator **cannot be the only indicator** used for answering of horizontal and measure specific CEQ such as:

- ▶ To what extent has the aid contributed to improving the competitiveness of the agricultural sector?
- ▶ To what extent have supported investments contributed to improving the diversification and development of the rural economy?
- ▶ To what extent have supported investments contributed to improving the quality of life in rural areas?

For example, the measurement of sectoral competitiveness requires using additional indicators, e.g.

- ▶ Share (%) of a given sector/group of farms in total output;
- ▶ Share (%) of a given sector/group of farms in total generated value added;
- ▶ Share (%) of a given sector/group of farms in sold production;
- ▶ Share (%) of a given sector/group of farms in exports

Furthermore, the assessment of the impact of a given programme on the quality of life in rural areas or the measurement of a contribution of a given programme to development of the rural economy requires using others, **synthetic indicators**, e.g. **Rural Development Index** (see: Michalek, 2008, 2009b).

Example

In (Michalek, 2009) the **overall net impact** of EU RD programmes on rural regions (aggregated effects of a given RD programme at regional levels) are estimated using the **Rural Development Index (RDI)** – a proxy describing the **overall quality of life** in individual rural areas. The weights of economic, social and environmental domains entering the RDI index (composite indicator) are derived empirically from the econometrically estimated intra- and inter-regional migration function. The impacts of individual RD measures are analysed by means of a counterfactual analysis by applying combination of the Propensity Score Matching (PSM) and difference-in-differences (DID) methods using the RDI Index, unemployment ratios, etc. as impact indicators. Given information on regional intensity to programme exposure (financial input flows by regions) the *overall* impact of obtained support via a given RD programme is estimated by means of a dose-response function and derivative dose-response function within the framework of a generalized propensity score matching (GPS). Above methodologies are empirically applied to evaluation of the impact of the SAPARD programme in Poland and Slovakia in years

2002-2005 at NUTS-4 level. Results show a full applicability of proposed approach to the measurement of the impact of rural development and structural programmes.

Obviously abundant micro-economic data and advanced evaluation methodologies are crucial elements in good quantitative analysis of programme effects. Additionally, it is recommendable to complement quantitative estimates of NAGVA-PPS with some **qualitative information**, e.g. by carrying out additional surveys focussing on displacement issues (labour, capital, etc.), or interviewing enterprises (programme non-beneficiaries) that feel to be affected by a given programme in order to find out a correct reference for a quantitative analysis.

4.2 Impact Indicator 2: Employment Creation

4.2.1 The CMEF Guidelines

The CMEF Handbook suggests measuring employment effects in *“Full Time Equivalent (FTE) jobs created expressed as the number of additional jobs created directly in supported projects and indirectly in the programme area that can be attributed to the intervention once double counting, deadweight, displacement and multiplier effects have been taken into account. Normally, a FTE is considered to last for at least 10 years.”*

FTE is the number of full-time equivalent jobs, defined as total hours worked divided by average annual hours worked in full-time jobs (UN SNA 1993). The average annual hours worked in a full-time job vary by Member States and data bases.⁶⁶ The CMEF (Guidance note J) further requires the employment indicator to be broken down by agriculture, forestry and non-agricultural activities, by gender and by age under/over 25.

The impact indicator employment creation contributes to the following horizontal evaluation question (Guidance note B):

- ▶ *“To what extent has the programme contributed to the realisation of Community priorities in relation to the renewed Lisbon strategy for growth and jobs with respect to the creation of employment opportunities?”*

Axis III also contains measure-related questions regarding additional employment opportunities due to diversification, business creation and tourism (measures 311-313):

- ▶ *“To what extent have supported investments promoted additional employment opportunities for farm households outside the agricultural sector?”*
- ▶ *“To what extent has the support promoted additional employment opportunities in rural areas?”*

The impact indicator corresponds to the result indicator 8 “gross increase of jobs”. The result indicator measures the number of jobs created (head count, not created FTE) and does not account for possible displacement and multiplier effects. Due to practical problems with data collection (see section data requirement), the result indicator should not on its own be used for impact evaluation. Larger

⁶⁶ See OECD Stat Extracts for average annual hours worked in different Member States:
<http://stats.oecd.org/Index.aspx?DatasetCode=ANHRS>. The exact figure needs to be checked in the database used.

employment impacts of RD measures are most likely to occur in non-agricultural sectors. Thus, the baseline indicator 28 “employment in non-agricultural sectors” should be broken down by the most relevant non-agricultural sectors in rural areas.

The CMEF suggests to analyse employment effects at the level of beneficiaries (micro level) and to estimate the contribution of these micro level effects to the general (baseline) trend at programme area level (macro level). This indicates a need for evaluation methods applicable at micro and macro levels.

4.2.2 Key challenges with regard to measurement and interpretation

The main methodological challenges of a comprehensive impact evaluation are to:

- ▶ Provide empirical evidence of a true **cause-and-effect link between** the observed indicators and the rural development (RD) programme; determining the counterfactual is widely considered as a core of each evaluation design. However, a counterfactual performance of the same regions, programme areas, etc., that participated in the programme cannot be directly observed (it is physically impossible to measure outcomes for a given unit of analysis (e.g. a region), which is participating in a programme and not participating at the same time). Therefore, the counterfactual has to be “constructed”;
- ▶ Consider and embrace all possible programme direct and indirect effects (positive and negative). This should be done by taking into consideration changes in all relevant impact indicators that were caused by the programme which implies deriving an appropriate baseline.
- ▶ Disentangle the effects of a programme support from effects of other exogenous intervening factors that may have also influenced a given “impact indicator” calculated at the regional/macro level. These influences may stem from other programmes, e.g. Structural Funds, or from programme-independent causes. Measuring the net effects means to subtract the changes which would have occurred in the absence of the public intervention from the gross effects also taking into account deadweight, leverage, displacement, substitution and multiplier effects;
- ▶ While certain impacts can be observed among direct beneficiaries (e.g. turnover generated for the suppliers of assisted firms), others can only be observed at macroeconomic or macro-social level (e.g. improvement of the image of the assisted region). Often, identifying impacts at the micro level seems to be easier than identifying overall impacts (DG Agri 2004).

Therefore the major challenge will be to find the optimal point of intersection between the costs of impact measurement and the quality standards necessary to get closer to the true impact; this will directly impact the requirements for data availability and quality. Quantitative impact assessments are often not realized due to (i) the diversity of RD support; (ii) the way in which monitoring systems are set up; (iii) partly also the relative small scale of RD schemes; and iv) doubts over “good value for money”. However, as long as secondary data are available to implement a quantitative analysis, it should be done because cost arguments should then weigh less than the potential gains from less biased assessments.

Furthermore, assessing employment effects in the context of policy evaluation should comprise an analysis of macro welfare effects. The question is here whether RD support contributes to a more efficient inter-sectoral allocation of labour.

4.2.3 Recommendable methods of measurement

4.2.3.1 General considerations

The quantification of employment impacts should be approached from two sides: At micro level, the mechanism of how RD support contributes to the creation of jobs is analysed. This allows quantifying gross employment effects accounting for deadweight, but not for displacement and multiplier effects. As the latter two effects occur in the (closer or further) surrounding of beneficiaries, the scope of analysis needs to be broadened to the macro level (NUTS 2, NUTS 3). Net employment effects and their contribution to general employment trends are to be evaluated there. The assumption is here that displacement and multiplier effects occur in the same (NUTS 2 or NUTS 3) region where the supported RD measure is allocated. The methods proposed in the following correspond to the micro and macro level of impact evaluation.

Micro level analysis of employment effects can be done by using control group comparisons of beneficiaries and non-beneficiaries. A crucial question of control group comparisons is how to find non-beneficiaries that have characteristics similar to beneficiaries. One option is to stratify both groups along their characteristics (firm size, educational level, etc.). With small populations of beneficiaries and thus small sample sizes one frequently ends up in subsamples that are too small to apply statistical tests. The idea of matching is to pool “twins” of beneficiaries and non-beneficiaries into one sample and compare the means of both groups. Among different matching methods, Propensity Score Matching (PSM) is the most popular one. Beneficiaries and non-beneficiaries are matched on the basis of the Propensity Score – an aggregated, univariate measure of multiple individual characteristics.

An often less preferable alternative for micro level analysis is a standard regression model with an Ordinary Least Squares (OLS) estimator. PSM is often preferable over OLS because the former does not impose linear relationships between variables or a homogeneous additive treatment effect. Thus, if PSM is implemented well, it will likely result in less biased estimates in comparison with OLS. PSM is generally more robust. On the other hand, if OLS is specified correctly (with a set of controls; if appropriate, includes higher order and interaction terms), it is more efficient than PSM, i.e. smaller sample sizes suffice for the estimation. Admittedly, this is not very likely in such a complex case like rural development policies. A limiting criterion for applying PSM is its high data demand. If all data have to be collected by surveys, the costs of applying PSM increase significantly. This is why secondary data is the more likely source for applying PSM.

Both methods (PSM and OLS) can be implemented with cross-sectional data observed at a single point in time (with-without comparison) and longitudinal data with observations of beneficiaries and non-beneficiaries before and after RD assistance (Difference-In-Difference comparison – DID). A DID comparison using longitudinal data is much more preferable because characteristics of beneficiaries and non-beneficiaries prior the RD assistance and unobserved factors are largely taken into account, which can lessen selection bias considerably.

The estimation of gross effects on micro level based on cross-sectional data⁶⁷ includes the following steps:

4.2.3.2 Propensity Score Matching

- ▶ Collect a representative sample of beneficiaries and non-beneficiaries in a survey, or more preferably from secondary data. The sample size of beneficiaries should exceed 50 observations. The sample size of non-beneficiaries should, at minimum, exceed the number of beneficiary observations by a factor 4 to 10 (200 to 500 observations). The more different beneficiaries and non-beneficiaries are, the larger should be the sample size of non-beneficiaries.
- ▶ Estimate the propensity score using logit or probit analysis, where the dependent variable is the observed decision to participate (= beneficiary) or not to participate (= non-beneficiary) in RD measures and the independent variables are individual (or communal) characteristics of beneficiaries and non-beneficiaries that influence the decision (not) to participate in a RD measure.
- ▶ Assign beneficiaries and non-beneficiaries with identical propensity scores to each other.
- ▶ Check whether mean differences in characteristics between beneficiaries and non-beneficiaries have been removed through matching. Use the T-Test. If mean differences between both groups are not significant, proceed further. Otherwise modify the specified logit model and include those variables (or higher order terms of them) as explanatory variables whose mean differences had not been removed through matching before. For further guidance see also Caliendo and Kopeinig (2000).
- ▶ Calculate the mean difference of the number of jobs (in FTE) between beneficiaries and non-beneficiaries, which is the gross employment effect.

4.2.3.3 Standard regression model

- ▶ Collect a representative sample of beneficiaries and non-beneficiaries in a survey, or from secondary data. As a rule of thumb, the sample size of beneficiaries and non-beneficiaries should, at minimum, exceed 100 observations or more (depending on the number of regressors used).
- ▶ Specify a regression model (OLS) where the dependent variable is the number of jobs (in FTE) observed for beneficiaries and non-beneficiaries. Independent variables include a binary (0 = non-beneficiary, 1 = beneficiary) or discrete (0 = non-beneficiary, 1 = beneficiary class 1, 2 = beneficiary class 2)⁶⁸ indicator of RD participation and individual (or communal) characteristics of beneficiaries and non-beneficiaries.
- ▶ Predict the number of jobs of beneficiaries with and without participation to RD measures. Calculate the mean difference between the two parameters (the gross employment effect).

⁶⁷ The methods described in the following can easily be adapted to longitudinal data. Practical examples for the estimation of gross effects based on longitudinal data (preferable over cross-sectional methods) are given in Box 1 (PSM) and in Box 2 (OLS with panel data).

⁶⁸ A continuous indicator of RD participation (e.g. amount of payments received) may lead to biased estimates because linearity is assumed between employment and the amount of payments (which is zero for all non-beneficiaries and above zero for beneficiaries).

Box 1 Real case example for a PSM combined with DID using data on farm level

Title and author: Employment effects of the less favoured area (LFA) scheme and agri-environment (AE) programmes in Hessen, Germany (Pufahl, 2008, only in German). A similar study is available in English (Pufahl and Weiss, 2009).

Background: The study was conducted for the ex-post evaluation of the RD plan (2000-2006) in Hessen/DE. The impact of the LFA and AE scheme on on-farm employment (calculated in FTE and FTE per 100 hectare farmland) was analysed at the level of individual farms.

Data: The analysis is based upon bookkeeping data (similar to FADN) of approximately 450 farms in Hessen. Bookkeeping data comprise information on farm, farmer and farm household characteristics as well as on the participation in the LFA and AE scheme (amount of LFA/AE payments received per farm/year). For each of the 450 farms, data from 2000-2005 is available. During that time 89 (107) farms participated to LFA (AE programs). *Beneficiaries of LFA* (LFA=1) are defined as those farms that have received a positive amount of LFA payments in all six years (2000-2005). *Beneficiaries of AE* (AE=1) are farms that have not received any AE payments in the base year (2000) and have received AE payments from at least 2002 continuously up to 2005. *Non-beneficiaries* (LFA=0, AE=0) are farms that did never receive LFA or AE payments from 2000-2005. Farm level data were supplemented by regional data on Nuts 3 level (e.g. unemployment rate, industrial wages). The table below shows the structure of the data set used.

Data structure

Farm	AEP	LFA	Farm FTE_00	Farm FTE_05	Rate of change	Farm size_00	other variables
1	1	1	1.3	1.2	-0.1	70	...
2	0	0	2.5	2.5	0	50	...
3	1	0	0.7	0.7	0	20	...
4	0	1	1.7	1.9	0.2	130	...
...

Note: Hypothetical figures.

Method: Employment effects of LFA and AE schemes are analysed by applying a control group comparison (DID) combined with Propensity Score Matching. A greedy matching estimator employing 1:1 matching without replacement was used to assign beneficiaries and non-beneficiaries to each other (Parson, 2001).⁶⁹

Results: In a first step, the propensity score is estimated using logistic regression. The dependent variable is the participation status of each farm. Explanatory variables included farm specific characteristics (e.g. farm size, share of grassland, expenditures for pesticides, livestock densities, farm profit, farm sales, age and education of farm operator) and regional characteristics (unemployment rate, real estate prices, land rents, industry wages, etc.) in the base year 2000. The propensity score is the conditional probability of each farm to be a beneficiary and takes on values between 0 and 1.

The selection of relevant explanatory variables was guided by prior information about what factors drive participation to LFA and or AE schemes. For instance, AE beneficiaries (at the average) manage their land (prior to participation) less intensively than non-beneficiaries. Thus, indicators proxying land use intensity (e.g. livestock density, share of extensive crops/grassland, expenditures for pesticides and fertilizer) are considered. Participation to LFA depends on whether a farm is situated inside or outside a disadvantaged area. Thus, criteria used to delineate the eligible area (e.g. soil and climate conditions, topography) supplemented by farm characteristics replicate factors influencing the decision to participate. Furthermore, using FTE (FTE per 100 hectare) as an explanatory variable ensures that both groups have a similar level of on-farm employment in the base year.

Beneficiaries and non-beneficiaries with similar propensity scores are assigned to each other. Only those 68 beneficiaries for which non-beneficiaries with similar propensity scores are available are used for further analysis. 39 beneficiaries had to be excluded from the analysis because no similar non-beneficiaries were available (because the pool of non-beneficiaries was too small). This violates the common support assumption⁷⁰ and may lead to biased results. After matching, it needs to be checked whether farm characteristics of beneficiaries and non-beneficiaries are on average similar (see table below). Differences in means are tested using a T-Test.

Means of selected characteristics before and after matching (AE programme)

Characteristic in base year (2000)	Unit	Before matching		After matching	
		Beneficiaries	Non-beneficiaries	Beneficiaries	Non-beneficiaries
Farmland	hectare	71.29	45.67 *	59.01	50.80
On-farm labour	FTE	1.49	1.32 *	1.44	1.39
On-farm labour/100 hectares	FTE	2.09	2.89 *	2.44	2.28
Farm income	EUR 1,000	22.91	16.02 *	18.89	16.15
Number of observations		107	237	68	68

Note: Asterisks indicate significant differences in means between beneficiaries and non-beneficiaries in the base year 2000 at < 5% level as measured by the T-Test.

The last step involves the calculation of the rate of change of on-farm employment between 2000 and 2005. Mean rates of change are computed for beneficiaries and non-beneficiaries (see columns (1) and (2) in the table below). The Average Treatment effect on the Treated (ATT) is the difference between the rates of change of both groups ((1)-(2)). The T-Test is used to detect whether observed differences in means are significant.

Employment impact of the LFA and the AE scheme in Hessen, Germany (2000–2005)

Indicator of interest	Unit	Rate of change Beneficiaries	Rate of change Non-beneficiaries	Average Treatment effect on the Treated (ATT)	
		(1)	(2)	(1)-(2)	
Less favoured area scheme		(n=54)	(n=54)		
On-farm labour	FTE	-0.03	-0.14	0.11	*
On-farm labour/100 hectares	FTE	-0.36	-0.15	-0.21	

⁶⁹ Parson, L. S. (2001): Reducing bias in a propensity score matched-pair sample using greedy matching techniques. Proceedings of the 26th Meeting of the SAS International User Group (SUGI 26): www2.sas.com/proceedings/sugi26/p214-26.pdf. Different user written matching estimators are available for all major statistical programmes, e.g. from Leuven and Sianesi (2003) for Stata.

⁷⁰ The assumption of common support states, that there should be a similar non-beneficiary for each beneficiary.

Agri-environment scheme		(n=68)	(n=68)		
On-farm labour	FTE	-0.16	0.00	-0.17	***
On-farm labour/100 hectares	FTE	-0.42	0.02	-0.44	***

Note: *** (**, *): significant at the 1% (5%, 10%) level.

Beneficiaries of LFA show almost no change in total on-farm employment from 2000-2005 while non-beneficiaries reduced total on-farm employment by 0.14 FTE per farm at the average. The difference between both groups results in an average effect for beneficiaries (ATT) of +0.11 FTE and is significant at a 10 per cent level. Against the general decrease in agricultural employment, LFA helped to maintain labour in agriculture. The opposite is true for AE programs, because support AE measures in Hessen predominantly support (labour) extensive production methods. This results in an average reduction of agricultural labour force of 0.17 FTE per farm (significant at 1% level). On farm labour per hectare also decreases strongly under AE participation (-0.44 FTE per farm, significant at 1% level) because the amount of farmed land is extended significantly under AE programs (see Pufahl and Weiss, 2009).

4.2.3.4 Assessing employment effects at macro level

The state of the art for analysing macro-impacts of employment policies are econometric panel models that require longitudinal data of at least 2 (fixed-effect models) or 3 years (GMM models). Fixed-effects models are estimated by OLS applied to demeaned (or within-transformed) data. Time demeaned data can be obtained by subtracting the annual deviation of a variable's overall mean. GMM models usually employ first-differenced data and internal instruments to control endogeneity. Both models are covered in standard econometric textbooks like Wooldridge (2002). Empirical examples of both models can be found in Box 2 and in Blien et al. (2005). Equivalent to the micro level, an advantage of panel techniques over cross-section approaches is that they can control for unobserved individual heterogeneity that is constant over time. Because time constant factors are controlled for, the indicator of programme participation (preferably the amount of RD payments in Euro) for one region need to vary over time.⁷¹ Spatial relations can be taken into account by including rural development expenditures of neighbouring regions as additional variable into a standard panel model. Net employment effects can be computed from model estimates, comparing the projected outcome with and without RD support.

Less practical alternatives for macro level impact evaluations are econometric cross-section models (like at the micro level), linear programming, general equilibrium (CGE) models and alike. Employment effects can basically be modelled as part of encompassing simulation models (CGE or structural economic models) like for example HERMIN, REMI, E3ME but because of the effort necessary to implement them, these might be seen to be more appropriate for larger Structural Funds interventions, and not for single programmes under the RDR unless such a model is readily available including the original authors doing the work. The HERMIN model (Bradley, et al. 2003) is a static CGE macro model developed for the ex-ante impact evaluation of investments supported by the Structural Funds. The REMI Policy Insight model (Treyz and Treyz 2004) is a dynamic regional economic forecasting and policy analysis model integrating input-output analysis, CGE modules, econometric equations using time-series data, and productivity and competitiveness effects due to agglomeration and clustering, which are based on new economic geography. The E3ME model⁷² is a dynamic econometric model that uses also some of the techniques used in CGEs with built-in feedbacks between the economy, population, demand/supply of energy and environmental emissions. Experiences show that the numerous assumptions made in such complex

⁷¹ Regions that never benefited from RD support or always received an identical amount of RD support over different years are excluded from the estimation of the treatment effect in fixed-effect models.

⁷² http://www.camecon.com/suite_economic_models/e3me.htm#

models are hard to elicit and revise by evaluators. These regional models partly also fail to adequately reflect regional specificities.

Box 2 Real case example of panel data analysis of employment effects at macro level

Authors and title: Petrick and Zier (2009): Employment impacts of the Common Agricultural Policy in Eastern Germany – A regional panel data approach.

Background: The objective of the study is to quantify employment effects of the Common Agricultural Policy (CAP) in the primary sector (including agriculture, forestry and fisheries) in three East German States (Brandenburg, Sachsen, and Sachsen-Anhalt). Previous studies examined the impact of single agricultural policy instruments only or treated CAP interventions as being one homogenous intervention. In this study, Petrick and Zier (2009) analyse the impact of different CAP instruments (e.g. direct area payments, AE payments, investment aids, etc.) on employment.

Data: A panel data set with $i=69$ NUTS 3 regions and $t=6$ years (2000 to 2005) is used. For each region i , six year-observations are available (balanced, non-rotating panel). Information about CAP support by instrument, region i and year t was made available by the paying agencies of the three German Federal States. Further control variables (e.g. population density) are taken from official statistics.

Data structure

Region	Year	Employ_1 sector	Direct area payments	other variables
1	2000	992	200,000	...
1	2001	985	210,000	...
1	2002	970	211,000	...
1	2003	960	213,000	...
1	2004	954	217,000	...
1	2005	946	220,000	...
2	2000	1,345	90,000	...
2	2001	1,340	80,000	...
...

Note: Hypothetical figures.

Method: The authors use three different panel models: A pooled OLS model (model A), an OLS model with regional effects (model B) and an OLS model with regional and year effects (model C). Model B and C correspond to the Least Square Dummy Variable Model (LSDV). Identical results can be obtained from a fixed-effect model (implemented in all standard statistics software packages). Model A is the most simple one and has the form

$$y_{it} = \alpha + \beta x_{it} + \delta d_{it} + \varepsilon_{it}, \text{ (Model A)}$$

Where y_{it} represents the number of employed persons in the first sector of region i at time t , α is a constant term, x_{it} comprises all variables that, besides CAP payments, may have an influence on primary sector employment (e.g.

wages) and d_{it} represents payments by different CAP instruments of interest. ε_{it} is the identically and independently distributed (i.i.d.) error term. The coefficients β and δ are to be estimated and represent the influence of x_{it} and d_{it} on employment. In Model A, the region and year specific structure of the data set is not taken into account; all observations are “pooled” into one model. Regional and year specific factors (e.g. farm structure, economic shocks) that influence y_{it} but are not included in x_{it} are not controlled. These unobserved factors are likely to cause biased estimates of the employment impact of CAP instruments. Model B allows for a regional differentiation of the model.

It includes 68 regional dummy variables (represented by η_i) taking either the value 0 or 1.

$$y_{it} = \alpha + \beta x_{it} + \delta d_{it} + \eta_i + \varepsilon_{it}, \text{ (Model B)}$$

The influence of time constant, region specific factors is captured by these dummy variables. Model C controls for region and year specific factors:

$$y_{it} = \alpha + \beta x_{it} + \delta d_{it} + \eta_i + \mu_t + \varepsilon_{it}, \text{ (Model C)}$$

where μ_t represents the effect of year specific dummy variables. The methodological considerations suggest that the constituency of the estimates should increase from model A to model C.

Results: Table below represents the estimation results for model A, B and C. Consistent with our expectations, the explanatory power of the models increase from model A to model C (see adjusted R^2). Same holds for the estimated policy impact of CAP instruments on primary sector employment. In model A, the impact of direct area payments on employment is significantly positive and becomes negative in model B and C since the influence of regional and year specific factors are controlled here. In model C, an increase in direct area payments by one million

Euros is associated with a decrease of employment by 32.7 FTE. This figure represents the employment effect of direct area payments averaged over all regions and years, holding all other factors constant.

Regression estimates: policy impacts on employment in agriculture

Dependent variable Employment 1 st sector	Unit	Pooled OLS model		OLS with regional effects		OLS with regional and year effects	
		(A)		(B)		(C)	
Explanatory variables		Coefficient		Coefficient		Coefficient	
Direct livestock payments	MEUR	22.7		-0.9		-13.0	
Direct area payments	MEUR	52.7	***	-57.9	***	-32.7	***
Agri-environmental scheme	MEUR	67.4	***	12.8	**	12.0	*
Compensatory allowance	MEUR	299.9	***	-8.8		4.2	
Investment aid	MEUR	290.3	***	4.1		9.3	
Processing & marketing support	MEUR	2.5		-14.1	***	-12.7	**
Rural development measures	MEUR	-17.1	*	1.4		4.2	
Population density	Person/sq km	0.1		-0.9		-1.7	**
Average yearly wages	MEUR	-0.1		-0.1	***	-0.1	***
Dummy variables							
Region		no		Yes		yes	
Year		no		No		yes	
Number of observations	N	414		414		414	
Number of years	t	6		6		6	
Number of regions	i	69		69		69	
Adjusted R ²		0.756		0.977		0.978	

Notes: All models include a fixed intercept. *** (**, *): significant at the 1% (5%, 10%) levels.

Unfortunately, the authors did not predict the employment level for regions with and without a specific policy support. The difference of these two parameters is the mean impact of the respective policy on beneficiaries.

A potential criticism of the empirical model in the table above is that only two variables (population density and yearly wages) are included as controls. Admittedly, these two variables cover quite a large terrain of the domain "rurality". However, it can still be envisaged that additional factors like agricultural structures, natural production conditions and non-agricultural economic developments influence agricultural employment and the probability to receive CAP support. Leaving these factors out will probably bias the results. Which additional or other control variables should be included can be learned from regional/national studies or own assumptions based on data availability and statistical significance.

The above approach can readily be implemented for the agriculture-related measures of Axis I and II. For the really econometrically minded evaluator, one could also envisage an application of the generalised PSM approach to employment as already suggested for the GVA indicator earlier in this paper.

Measures of Axis III and IV are insofar of a different nature as they aim to affect socio-economic regional structures indirectly by creating networks, improving infrastructure, etc. and because questions arise on how much and which employment is created (outside of agriculture) through diversification, business creation in general, and tourism where availability of secondary data is often lower.

4.2.3.5 Getting to know how the policy works

A mix of qualitative and quantitative methods could be applied to cover possible effects on employment outside of agriculture and/or to get insights not only on the magnitude of the effect (quantitative method) but also on how RD policies affect individuals, communities or regions (often qualitative

methods). The latter is done to identify factors supporting and hampering the successful implementation of RD policies.

One option to link qualitative and quantitative approaches in a triangulate fashion is to extend the PSM approach to the community level⁷³ to compare selected beneficiary and non-beneficiary communities (with-without, preferably combined with DID). One can envisage a PSM application embedded in a wider (matched) case study design using programme logic models, in which pattern matching (for rival explanations, of which the intervention logic can be one) is combined with a descriptive time series analysis (Yin 2008) and other qualitative methods.⁷⁴ This combination would not only shed light on whether a policy worked and how much the effect is, but also gives indications on why and how measure implementations do (or do not) work.

The PSM approach at community level needs to be amended from the above example (box 1) insofar as

- ▶ the units of analysis are not farms but communities that have received (beneficiaries) or have not received (non-beneficiaries) Axis III/IV measure support;
- ▶ the treatment (yes=1, no=0) can be defined as receiving support via Axis III/IV for single measures (311, 312, 313) or the aggregated support via Axis III/IV;
- ▶ the outcome variable is the number of FTE in the community after the RD-support (cross-sectional), or, preferably, before and after the RD-support (DiD); and
- ▶ the estimation of the propensity scores should be based upon factors that influence community participation in Axis III/IV measures (ability to co-finance, administrative capacities, high unemployment rates, etc.) and factors that may influence the development of FTE in the community (population, population density, economic structure, infrastructure, etc.).

A community based approach requires a wide range of data (factors influencing uptake of RD measures and FTE development, employment data, financial data of RD support) that are available at community level (LAU 1, formerly NUTS 5). The critical point of applying PSM at community level will be to ensure minimum sample sizes as suggested above.

4.2.4 Data requirements and collection

Data at micro level should be collected from beneficiaries and non-beneficiaries on the basis of secondary data and/or own surveys. FADN data provide data of beneficiaries and non-beneficiaries for agricultural RD measures as farm investment support (measure 121), LFA schemes (measures 211, 212) and AE programmes (measures 214, 215). The suitability of FADN data for RD evaluation could be enhanced if different types of RD support would be further disaggregated (e.g. differentiation of various AEM or investment types).

⁷³ The local community level is suggested because this is what Axis III and IV measures focus on.

⁷⁴ Other methods are shortly described in the Appendix of Leeuw and Vaessen (2009).

Outside the agricultural sector, the availability of secondary data becomes scarce. Various national statistics on households (labour force surveys⁷⁵) and enterprises are available, but availability varies strongly among Member States. In many cases, own surveys will be the only source of information for the evaluation of non-agricultural RD activities.

Data availability determines the applicability of micro level evaluation methods. PSM is very data demanding, because more observations are needed for non-beneficiaries than for beneficiaries. Furthermore, matched sample sizes should be sufficiently large to allow statistical testing. In terms of data requirements, OLS is less demanding than PSM. The suggested minimum sample sizes for PSM and OLS are mentioned above. The collection of primary longitudinal data requires repeated surveys which double the costs of data collection.

For various reasons the result indicator 8 “gross increase of jobs” cannot be used on its own for assessing the impact of RD support: First, the result indicator intends to count the number of created jobs and not the number of created FTE. Second, the result indicator is collected by the administration, which makes inconsistencies (due to unclear definitions and the involvement of many persons) and strategic answering likely.⁷⁶ Third, the result indicator frequently reflects “planned jobs” and not “created jobs”. The first problem could be tackled by including a result indicator “gross increase of jobs” calculated in FTE in the monitoring system, where full-time is defined with a certain amount of hours per week that varies between countries, half-time is 0.5, and part-time is normally assumed to be between 0.2 and 0.3 FTE.

Data for macro-level analysis are to be collected from official statistics. NUTS 3 (in some Member States even at LAU I level) might be the lowest possible macro level where employment effects of the RD support can be estimated (although this may vary by Member States). Relevant data stem from regional economic accounts which include data on employment (number of jobs or FTE, sometimes differentiated by sectors). Additional information on NUTS 3 (LAU I) level (e.g. wages, population size, firm structure, etc.) should be available from various official statistics. Further data on labour statistics are sometimes provided by the national labour offices. They typically count employed persons excluding self-employed and civil servants. The data distinguish between full-time, part-time and marginally employed persons. However, this data source does not help if the majority of employment effects of the RD support are expected to impact the number of self-employed persons (farmers, start-ups in diversified enterprise, etc.). Additional surveys or case studies will be necessary to get the necessary information.

As mentioned, the CMEF (Guidance note J) requires the employment indicator to be broken down by agriculture, forestry and non-agricultural activities, by gender and by age under/over 25. With respect to the practical estimation, the suggested differentiation is only applicable for micro-level evaluations where displacement and multiplier effects are not accounted for. At macro level, data allow a differentiation by sectors only.

Data on RD support should be gathered from the paying agencies at the micro level broken down by type of supported activity, type of supported beneficiary, regional/postal zip code of the supported

⁷⁵ Eurostat (2009): Labour force survey in the EU, candidate and EFTA countries. This publication describes the main characteristics of the national surveys. Eurostat also produces working papers on the topic:
http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-08-007/EN/KS-RA-08-007-EN.PDF

⁷⁶ This view is supported by experiences of evaluators with monitoring data.

project, public expenditures, etc. Aggregation of total expenditures to NUTS 3 or 2 levels is possible via regional codes and/or postal zip codes. Support provided by other instruments (1st pillar policies, Structural Funds, etc.) should also be taken into account, especially when fitting macro level models.

4.2.5 Interpretation and judgment issues

The estimation methods described reside on a number of assumptions which have to be critically checked for each single case (see e.g. Wooldridge, 2002). If all assumptions hold and there are no data problems, the quantified impact indicator(s) allow(s) the following statements: A positive net employment effect indicates whether the RD intervention helped to maintain jobs (against a general downward trend) or to create new jobs. The difference of gross employment effects (obtained from micro level analysis) and of net employment effects (obtained from macro level analysis) indicates the magnitude of deadweight, leverage, substitution, displacement and multiplier effects.

4.2.5.1 Possible hurdles for obtaining reliable results

However, there are difficulties in estimating the “true” employment effect for individual RD measures and the entire programme. These difficulties are listed here together with possible solutions to attenuate or circumvent them:

- ▶ **Time lag:** Employment effects may occur in short-term (e.g. as a consequence of rationalisation investments) or in mid- to long-term (e.g. as a consequence of educational, infrastructure or forestry investments). The evaluated period comprises 6 years (2007-2012) at maximum, because official statistics often become available only 1.5 years in retrospect. Long run effects of RD support could be addressed by considering projects, farms, communities in the former planning period (2000-2006). If the evaluator is restricted to analyse RD support from 2007 to 2013 there is only little option to address the time lag of RD policy impacts adequately. One option is to base the evaluation upon cases that are supported during the first 3 years of the planning period (2007, 2008 and 2009). This will not give reliable results for the mid-term evaluation, but can form the base for an ex-post evaluation. However, it should be taken into account that administrative support evolves slowly during the early programming period suggesting that only a relatively small number of projects are supported during that time.
- ▶ **Missing critical mass:** Effects of RD support are often too small to be separated from other unspecific factors (“white noise”) influencing employment. Especially non-agricultural RD support is “scattered” over the whole territory and channelled through a wide variety of RD measures. Obviously, a variety of instruments is useful to respond to different needs of regions, but it complicates applying quantitative evaluations. If this comes with overall low RD support, one could come to the conclusion, like in the Netherlands, that measurable macro effects may not to be expected for 2007-2013.
- ▶ **Effects in other regions:** The calculated net employment impact on NUTS 2 or NUTS 3 level considers only those substitution and multiplier effects that appear in the same (NUTS 2 or 3) region where the RD support was allocated. Effects outside the respective region (e.g. through the import of investment goods or the shift of product market shares through processing support) are not accounted for.
- ▶ **RD support is not mono-causal:** The creation of new jobs critically depends on how RD support is implemented. Focusing on effects only will shed no light on the conditions that

facilitate or hamper the creation of jobs. This “black box” can be illuminated by additional (qualitative and/or quantitative) methods which are based on theory.

- ▶ **Welfare effects:** When assessing the welfare effects of RD policies from a macro-economic perspective it has to be taken into account that in general, factor productivity in agriculture lags behind those of other sectors. This means that employment creation in agriculture should be evaluated against a wider conceptual background including goals like food security, territorial resilience and demographic balance.

4.2.5.2 Judgement issues

The results obtained from micro and macro level analysis can be used to answer horizontal and measure-specific evaluation questions.

Aggregating the employment effects estimated at macro level provides the quantitative part of the answer to the first horizontal evaluation question:

- ▶ “To what extent has the programme contributed to the realisation of the Community priority ‘creation of employment opportunities’?”

Further qualitative information in form of case studies etc. is needed if one wants to understand the mechanisms how RD support (on the side of beneficiaries, administration, etc.) hampers or facilitates the creation or retention of employment. For example, by matching similar case studies based on multiple sources of evidence, and investigating their development over time can substantially increase the validity and reliability of research designs (Yin 2008).

The employment indicator (and hence estimated effects) should be further disaggregated into non-agricultural sectors (construction, tourism, etc.) because most new jobs are expected to be created there.

The employment indicator sheds also light on the horizontal evaluation question:

- ▶ “To what extent has the programme design been successful in avoiding deadweight and/or displacement?”

The difference of gross and net employment effects indicates the magnitude of multiplier and displacement effects. The relevance of deadweight effects needs to be judged at micro level: If beneficiaries show a similar increase in jobs as comparable non-beneficiaries, the positive employment effect would have been occurred in the absence of the RD support, too.

The measure-related questions in Axis III regarding additional employment opportunities for farm households due to diversification (measure 311), business creation (measure 312) and tourism (measure 313) can be answered through the application of above-mentioned mix of quantitative and qualitative methods. Additionally, employment effects occurring due to Axis I (measures 121, 123) and Axis II measures (measures 111, 112, and 114) need to be taken into account and quantified by the evaluator. The evaluation would give an incomplete picture if unintended and negative employment effects of RD policies are not reported.

All socio-economic impact indicators (economic growth, employment creation and labour productivity) should be interpreted in a common context. For instance, rising total factor productivity (labour, capital, land) may explain why jobs have been lost in the agricultural sector.

4.3 Impact Indicator 3: Labour Productivity

4.3.1 The CMEF guidelines

The CMEF Guidance note J states that labour productivity should be measured as the *“Change in Gross Value Added per Full Time Equivalent (GVA/FTE).”* The indicator is defined as the *“Change in labour productivity is the Change in Gross Value Added per full time equivalent (GVA/FTE) in beneficiary population targeted by interventions and indirectly in the programme area that can be attributed to the intervention once double counting, deadweight, displacement and multiplier effects have been taken into account.”*

GVA is defined as *“value of output less the value of intermediate consumption”*. For the definition of FTE see section 4.2.1. (previous chapter on Employment Creation) The indicator ‘labour productivity’ provides an indication of how much value added one FTE is able to generate.

The CMEF established common evaluation questions displayed in Guidance Note B. The most relevant questions related to the impact indicator Labour Productivity are (relating measures in brackets):

- ▶ *“To what extent has the aid/support contributed to improving the competitiveness of the forestry holdings (122)/the agricultural and forestry holding through improvement of infrastructures (125)/the supported holdings” (Measure 131)?*
- ▶ *“To what extent have supported investments contributed to a better use of production factors on agricultural holdings? In particular, to what extent have supported investments facilitated the introduction of new technologies and innovation?” (Measure 121)*
- ▶ *“To what extent have supported investments contributed to maintain the economic performance of agricultural holdings through the restoration and/or preservation of the agricultural production potential?” (Measure 126)*

Some common evaluation questions have only an indirect relation to the impact indicator Labour Productivity as they are focusing on overall sector effects rather than on direct and indirect beneficiaries. However, findings at micro level can be used to support judgements upon overall sector effects

- ▶ *“To what extent have the actions related to training, information and diffusion of knowledge and innovative practises improved the labour productivity and/or other elements related to competitiveness in the agricultural, food and forestry sector?” (Measure 111)*
- ▶ *“To what extent has the aid/support contributed to improving the competitiveness of – the agricultural/forestry/food sector (Measures 112, 113, 114, 115, 121, 123, 124, 126, 132, 133)/the agricultural sector in the new member state?” (Measures 141, 142)*
- ▶ *“To what extent have supported investments contributed to improving the efficiency in the processing and marketing of agricultural and forestry products?” (Measure 123)*

4.3.2 Key challenges with regard to measurement and interpretation

4.3.2.1 Data availability and quality

Data availability is a general concern; information gaps vary from member state to member state, but in general food sector and forestry data are more difficult to obtain than agricultural data. Furthermore the quality of data such as bookkeeping figures provided by beneficiaries is often insufficient and requires reassessment.

4.3.2.2 Assessment of indirect effects

Indirect effects are difficult to identify; in order to gauge indirect effects it is necessary to establish wellfounded hypotheses on the result chains of the programmes and the relevant measures. The state of the arts of assessing impacts currently allows for measuring indirect effects occurring in the same sector. Identifying indirect effects on other sectors can be described, but there is no approach available to include these in the calculation of net effects.

4.3.2.3 Calculation of net effects

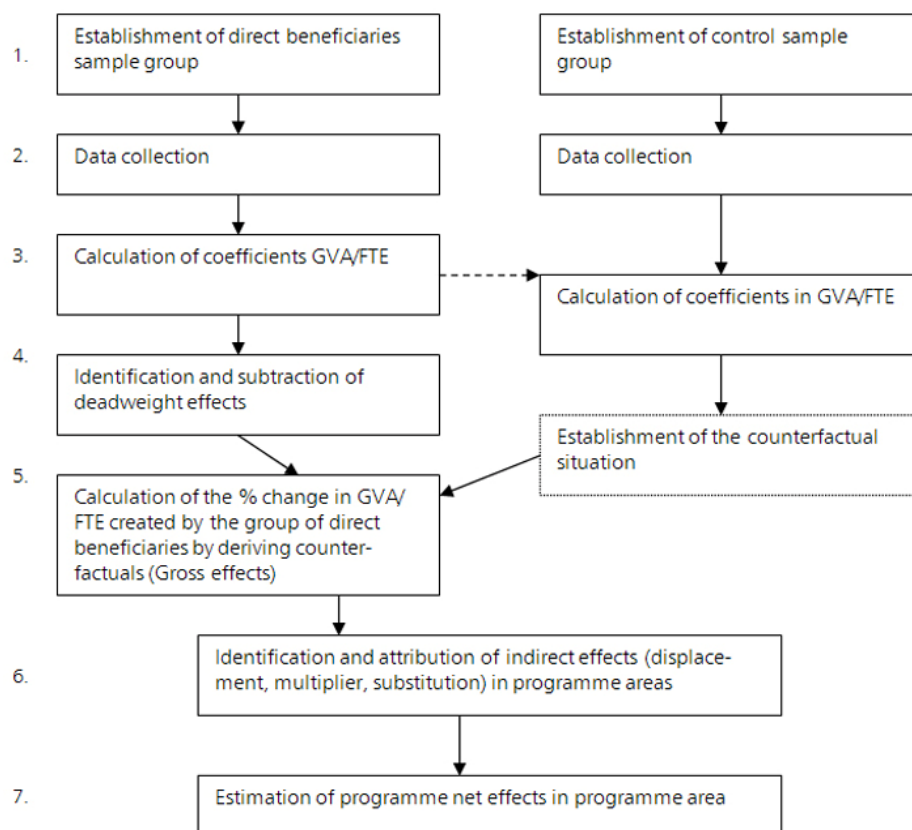
As mentioned above it is difficult to collect comprehensive data on indirect effects. Double counting and deadweight effects are occurring at the level of direct beneficiaries and can be accounted for; indirect effects like multiplication and displacement effects can only be estimated using macro level approaches.

4.3.3 Recommendable methods of measurement

Three baseline indicators (labour productivity in agriculture, in food industry, and in forestry) are related to this impact indicator. Labour productivity values for different sectors cannot be compared; they should therefore not be aggregated to a total value including all sectors concerned. Instead it is recommended to aggregate average values for GVA/FTE for each sector (agriculture, forestry, food) separately at programme level. In case of widely varying production systems, a disaggregation of the data is even necessary within the same sector.

The steps to measure the programme impact on labour productivity are depicted in the following figure.

Figure 4 Approach to establish gross and net programme effects at micro-level



Step 1: Identification of sample groups of direct beneficiaries, and comparison group

Programme authorities have to identify which beneficiaries to include in the data collection according to the implementation regulation of the programme. A thorough analysis of the result chain has to be made to identify the potential indirect beneficiaries. The micro level approach only allows accounting for indirect effects on labour productivity in the same sector and region. The establishment of the sample group has to account for similarities of measure implementation, measure objectives and the financial weight of specific measures. Furthermore, the possible aggregation of impacts of individual measures contributing towards specific objectives e.g. increasing human capital, or physical capital has to be considered. In order to establish the net effects, it is necessary to identify what would have happened without intervention (i.e. to establish the counterfactual), at best by a combined application of a sound matching method (Propensity Score Method) and the Difference-in-Difference Method, which are described in Chapter 3 above. The methodology for establishing the counterfactual (i.e. the establishment of a control group) needs to be chosen regarding the nature of support concerned; the availability of data, and cost-effectiveness of the chosen method.

Step 2: Data collection

Micro level data required for computing gross-value added and on full-time equivalents should be collected at household/holding level of beneficiaries and non-beneficiaries. FADN data for the agricultural sector can be used as a source for beneficiaries and non-beneficiaries information. For forestry and food sector data, other sources of data have to be identified; in most cases additional surveys are required (also see comments on data collection in section 4.2.2). Data need to be collected

separately for supported farm holdings, forestry owners and food companies, as well as for the sample groups described in Step 1. There is a need to improve statistical information on the forestry sector in many countries.

In order to enable the calculation of percent change in GVA/FTE, it is necessary to collect data of at least two points in time; before (t_0) and after (t_1) the support/programme. The most suitable time to collect significant data after participation in the programme is varying greatly depending on the nature of the support, esp. of the investment.

The following example from France explains how FADN data are combined with agricultural census data.

Real case example of data collection from existing statistical datasets

France: Combined Use of Quantitative, Mapping and Qualitative Analysis – Part 1: Making full use of existing statistical datasets. Claude Saint-Pierre (Tercia consultants)

Overall labour productivity can be obtained from national accounts of the agricultural sector, which are provided by the national statistical centre INSEE⁷⁷. The farm accountancy data network (FADN) allows the analysis of the indicator at a more detailed level. Gross-value added per full-time equivalent, the RDP labour productivity indicator, can easily be derived from FADN on an annual basis. The FADN sample (7,500 holdings of which 40 percent are RDP beneficiaries) is often seen as non-representative because it describes professional farm holdings. However RDP beneficiaries are overwhelmingly professional farms in mainland France. One main limitation in this sample is that diversification income is not tracked, and holdings with substantial diversification incomes are outside the sample.

The agricultural census takes place every 10 years and covers all holdings above 1 hectare. Inter-census farm structure surveys take place on years 3, 5 and 7 after a census and cover 10 percent of holdings. A 2007 survey is available, and a 2013 survey should be available. Census and inter-census surveys also track beneficiaries from the main RDP measures. They provide standard gross margin (derived from FADN regional average) and FTE. Standard gross product before subsidy will be provided instead of gross margin starting in 2010 due to decoupling of CAP subsidies. Examining both FADN and inter-census surveys provides a fair assessment of trends in productivity.

A similar approach is used in Italy to overcome some of the limitations of the FADN.

Real case example of data collection: use of FADN data

Italy: Data samples drawn from FADN survey – the Piemonte case. Daniela Storti (INEA)

In Piemonte the regional body in charge for RDP evaluation NUVAL developed an approach for data improvement using the FADN network data: samples from the FADN survey complement the data collected for the evaluation.

Representative samples are drawn from the FADN survey (called the satellite group), covering both direct beneficiary farms and a control group. The links between the FADN samples and the data collected for the evaluation are ensured through the collection of a common set of information at farm level. Representative samples are extracted from administrative archives. The FADN is then enlarged to include also the representative sample. Currently the FADN sample for Piemonte includes 1000 farms, while the specific evaluation satellite group includes over 300. The cost for additional data collection is 300€ per farm.

The access to the FADN data is facilitated with specific agreements of the regional managing authorities with INEA, the Italian National Institute for Agricultural Economics, who is responsible for the FADN survey.

In order to address the lack of data on food sector holdings a survey has been conducted in Germany during the period 2000-2006. The approach is illustrated in the box below.

⁷⁷ ADE consulting, Pollen conseil and Edater. Ex-ante evaluation of the 2007-2013 hexagonal RDP.

Real case example of data collection: inquiry sheet for food sector holdings

Germany: Inquiry Sheet for Code 123 (Improvement of processing and marketing of agricultural products). Dr. Antje Fitschen–Lischewski (vTI).

- **The inquiry sheet** (Excel sheet) was developed in the evaluation period 2000 – 2006 by the former FAL (now vTI), Institute of Market Analysis and Agricultural Trade Policy
- Collecting business indicators from investing companies
 - at the point of application of permit (**initial situation** indicators (t) and **planned** indicators **one year after completion** (t+1))
 - after completion of the investment (**real indicators one year after completion** (t+1))
- Excel sheets are transported into SAS, resulting in a data basis to answer different evaluation questions

Dimension "Production / Sales / financial figures "				
Processed goods	Unit	Year before application for approvement (t)	Planned for the year after completion (t+1)	
Processed goods Turnover	Euro			
Newly launched products in the company		Yes / No	Yes / No	
if yes, units	Units			
Turnover with new products	Euro			
Launch of a new product line in the company		Yes / No	Yes / No	
if yes, units	Units			
Turnover with new products	Euro			
Production of high quality produce with regard to EU-regulations (each with turnover in Euro)				
Organic products	Turnover	Euro		
Labels for regional origin (e.g. protected designation of origin)	Turnover	Euro		
Vintage wine	Turnover	Euro		
State regulations on food quality	Turnover	Euro		
Other regulations on quality	Turnover	Euro		
Employees on year base (FTE) (Please refer to the comment!)	Unit	Year before application for approvement (t)	Planned for the year after completion (t+1)	
Full time Equivalent	Number FTE			
thereof female	%			
Apprentices	Number			
Financial figures (Please read the comments on the items)	Unit	Year before application for approvement (t)	Planned for the year after completion (t+1)	
Total turnover (incl. other revenues)	Euro			
Increase or decrease of finished or unfinished supply	Euro			
Self produced structures and buildings (other booked internal activity)	Euro			
Consumption of (expense for) primary, auxiliary and operating products	Euro			
Expense for purchased trade goods	Euro			
Expense for purchased services	Euro			
Other inputs	Euro			
Gross value added (GVA)	Euro	0	0	
Personnel expense	Euro			
Total depreciation	Euro			
Total investment	Euro			
Proportion of R&D in total turnover of the company	%			
Proportion of advertising & promotion in total turnover of the company	%			

Step 3: Calculation of the % change in value added/full-time employment created

The calculation of the change of GVA/FTE is based on the definitions and prescribed components of the indicator in Eurostat.⁷⁸ The calculation of percent change is done through deducting the value of t_0 from t_1 and multiplying the value with 100. The value has to be calculated at single beneficiary level.

Step 4: Identification and deduction of deadweight effects

Deadweight effects may occur at direct beneficiaries; some prominent examples are provided below:

- ▶ If RD funding for investment support (measures 121, 122 and 123) is provided to those who are better off anyway, and who could have raised private funds;
- ▶ Implementing measure 132 ("support for food quality schemes") may cause deadweight if used in programme areas where farmers are already urged to join these schemes by their retailers or processors which means that at least a part of the costs incurred have already been internalised within supply chain relationships.

There are different options to identify deadweight effects; for example through questions to beneficiaries; profitability criteria (for investment support), financial criteria, or by comparison with farmers who make investments without support. The cases of identified deadweight have to be recorded and excluded from the calculation of the average % change of GVA/FTE.

Step 5: Calculation of the % change in GVA/FTE created by the group of direct beneficiaries by deriving appropriate counterfactual situations

The following practice example from France introduces an approach for the calculation of trends for beneficiary and non-beneficiary farms based on FADN data.

Real case example of comparing trends for specific target groups and areas

France: Combined Use of Quantitative, Mapping and Qualitative Analysis. Part 2: Comparing trends for specific target groups and areas. Claude Saint-Pierre (Tercia consultants)

Cemagref, the water and land management research centre, has carried out a comprehensive review of alternatives in the definition of counterfactual evidence and tested a method of matching pairs (holdings with comparable explanatory variables). Difficulties in identifying matching pairs, although expected, could not be fully resolved, particularly for the agri-environment scheme for extensive pastures.

FADN can be disaggregated into samples of RDP beneficiaries and non-beneficiaries with comparable features for single schemes. This led to identify higher productivity gains among beneficiaries of modernization plans in the ex-post evaluation. However the FADN sub-sample (60 beneficiaries) was too small to confirm significance of the differential with the control group.

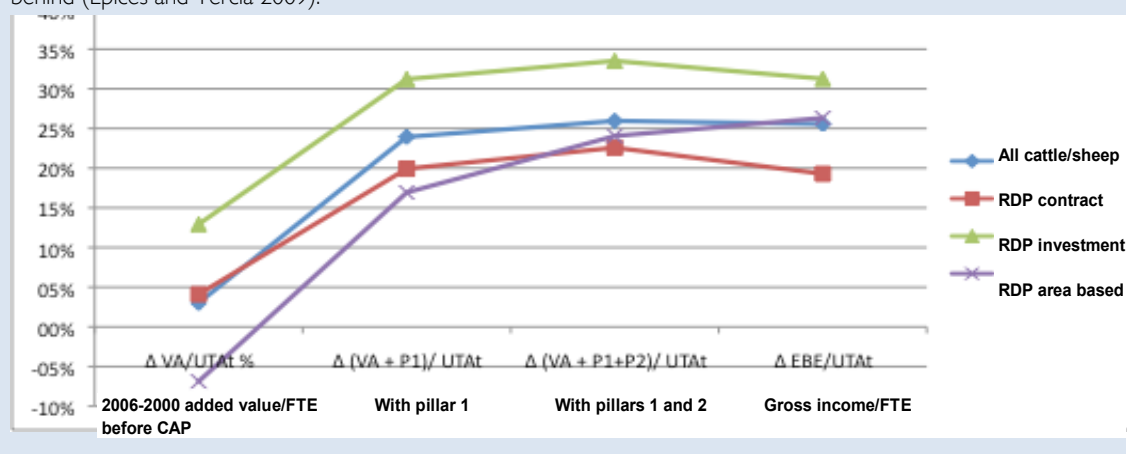
With well-defined target groups and target areas in France, an alternative is to analyze the indicator for a series of beneficiary types and/or area types. This implies the use of an "if/then" logical equation:

"If productivity has increased for groups 1, 2, ..., n (If productivity has increased for areas a, b, ..., m), THEN the RDP has contributed to productivity increases."

Using this approach, the ex-post evaluation provided evidence that, within grass-based livestock producers, productivity in beneficiaries of investment schemes did grow faster than average in 2000-2006. Beneficiaries of contract schemes were able to sustain productivity growth at average level while LFA measure beneficiaries fell

⁷⁸ According to EUROSTAT 'Gross value added (GVA) is defined as the value of all newly generated goods and services less the value of all goods and services consumed as intermediate consumption. The depreciation of fixed assets is not taken into account.

behind (Epices and Tercia 2009):



An innovative approach is the use of GIS in order to identify categories of beneficiaries and causal factors. The following practice example from France explains the approach and its usefulness.

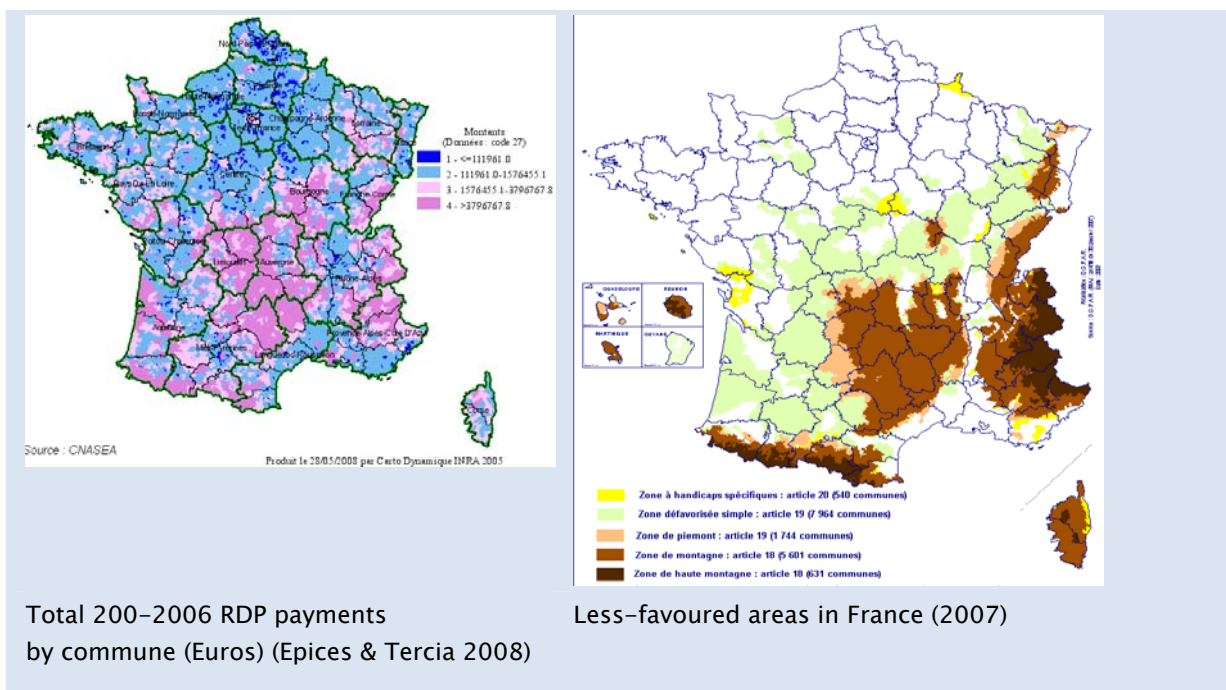
Real case example of GIS-based monitoring system for identification of beneficiary categories and causal factor

France: Combined Use of Quantitative, Mapping and Qualitative Analysis. Claude Saint-Pierre, (Tercia consultants)

Part 3: The GIS-based monitoring system, an “engine” to identify beneficiary categories and causal factors.

The 2003 mid-term evaluations saw the emergence of GIS as a tool for detailed analysis. INRA, the French national institute for agricultural research centre, set up at that time the rural development observatory (ODR) that provides dynamic mapping capacity for a broad range of variables. Maps are available down to commune level. With 36,000 communes in France, ODR provides very detailed pictures that can be correlated to variables with a well-known geographical distribution – a function that complements NUT 2 or 3 level maps. Correlations identified in a visual manner can be confirmed through quantitative analysis.

The 2003 maps have been instrumental in confirming how RDP measures in France are de facto targeted on the less-favoured areas and on livestock producers (Maps 1 and 2). This derives from three factors: (a) the mountain policy and the farm structure policy remain at the core of the RDP strategy, (b) the range of schemes in the RDP favour these producers and areas, and (c) a fair proportion of the larger holdings specialized in field crop production tend to be self-excluded from RDP measures which encourage a reduction of inputs.



Maps have also been used in the ex-post evaluations as an engine to identify linkages between variables. For example, the detailed geographical distribution of the two CAP pillar benefits was important for understanding how the second pillar compensates the productivity gap in RDP beneficiary holdings (map on the right). This in turn led to analyse the data at three stages: before CAP, after first pillar, after first and second pillars (Epices and Tercia 2008).

Step 6: Identification, valuation and attribution of macro level effects including displacement and multiplier effects

A macro level approach has to be applied to identify all indirect programme effects including displacement and multiplier effects. Two methods are currently available for macro level analysis: models (statistical and process-based) and qualitative methods. (see chapter 3 for further guidance)

The task to estimate displacement and multiplier effects is most challenging and requires using a combination of methods.. A calculation of net effects at micro level has many limitations as it can only include displacement and multiplication effects that are expressed in labour productivity limited to a specific sector and the observed region.

Socio-economic displacement effects possibly occur in response to any measure directly improving competitiveness of a business, sector or region, as other regions could lose their respective business advantages (measures 121, 123). Most displacement effects of axis I measures are likely to become obvious in the area of employment and income, and are not likely to be observable in labour productivity. A thorough reflection on the impact pathways (processes leading to change) is required to identify where displacement effects can occur and how to estimate these effects.

In order to be able to identify all displacement effects, a macro level analysis is necessary. Furthermore, this analysis has to make use of all three socio-economic impact indicators in combination for estimating the net programme effects.

Multiplier effects can be measured through the establishment of a sample group of indirect beneficiaries; they are added to the overall effects.

Step 7: Estimation of programme net effects in programme area

The final step in the process of measuring the impacts with regard to labour productivity of the three sectors is to assess the gross effects calculated at micro level data against the macro-level effects. Thus the identified gross effects for beneficiaries are compared to the observed general trends resulting from the macro level analysis. Labour productivity is also subject to cyclic fluctuations; this needs to be considered when interpreting the results.

In order to overcome the attribution gap between observed effects and total programme effects, qualitative methods (interviews, focus groups, surveys and case studies) are used in combination with quantitative analysis. The example from Belgium gives insights on this approach.

Real case example of theme-specific approach for labour productivity at micro-level

Belgium: Impact Assessment of Rural Development Programmes.... the Flemish experience. Katrien Van Dingenen (IDEA Consult). Theme-specific approach for labour productivity.

The very basis of the analysis on the themes of productivity/income was the structuring and use of a diverse set of complementary data and information sources, being:

- An expert-workshop to identify and validate the most important (potential/expected/identified) causal relationships (~gross effect) between the RDP-measures and the themes concerned:
 - Identification of theoretical/potential impact of specific (sub-)measures within the programme on productivity (all necessary circumstances such as sufficient budget & take-up, viability of farms...)
 - Estimation of the *level of* expected impact (5 category-level)
 - Validation of methodological assumptions and draft outcomes

This workshop was a combination of both plenary and thematic sessions (according to the axes of the current RDP), involving both academic staff & sector experts and public administration personnel from different departments.

- A widespread survey among a representative sample of Flemish farmers, with questions on:
 - Which investments they made, or which agro-environmental schemes they adopted, and
 - Whether they did it with or without RDP-support
 - [if support was obtained,] whether (to what extent) they would have done it without RDP support (~additionality)
 - [if no support was asked/obtained,] why this was the case
 - Which effects these investments had on the labour conditions, production techniques, productivity etc in their farm (using 5 categories, ranging from negative influence to a major positive influence)
 - What triggered their decision to do this investment/to follow a training/to adopt an agri-environmental scheme/to diversify their activities (only one of them being RDP support) (~contextual analysis ⇒ net effect)

This survey enabled the evaluators to get semi-quantitative & structured information on issues, and at a level of detail that is not present in any database or institution. This source can be seen as a more subjective source of information – but with high added value as it gives very focused input for the evaluation at micro/individual level.

- The accounting data in the FADN database for these (surveyed) farmers, with information on:
 - Employment & working hours per activity
 - Added value created
 - Output/turnover
 - Other farm-specific characteristics (type, ESU, geographical situation,...)

The evaluation team worked with a sample of dairy farmers for more detailed analysis within a uniform group of farmers – which could be divided in 3 groups: (i) farmers having received RDP-support, (ii) farmers with investments but not supported by a public authority and (ii) farmers without any major investments during the examination period.

- The accounting data in an IDEA-owned database called Belfirst – for the agrifood industry; with about the same information as for the FADN database
- The database of the Flemish Agricultural Investment Fund – indicating in a high level of detail the type of investment for each farmer that made use of this Fund
- The application dossiers for the agrifood measure (9.3.7 during RDP '00-'06)

The coupling of the accounting databases (FADN & Belfirst) allowed the evaluators to compare farms/companies

with RDP-support and others – to detect correlations between e.g. investments and support (*what is the importance of investment support and the actual investments made?*); and between investments (with and without support) and evolutions in added value created per FTE etc (*what is the micro level impact of these investments on the RDP-objectives?*).

These correlations at the level of farm investments could then be compared with the results of the survey. The survey also complements these data sources for the measures related to agri-environmental schemes and educational projects.

The result of this analysis was a **focused** but **cross-measure** and **micro level** assessment of potential and identified impacts of the Flemish RDP.

In France case studies have been used as a complementary method to provide in-depth insights and to confirm causal factors.

Real case example of local participatory case studies

France: Combined Use of Quantitative, Mapping and Qualitative Analysis. Part 4: Local participatory case studies: confirming causal factors. Claude Saint-Pierre, (Tercia consultants).

Case studies in evaluations are devoted to providing in-depth insights on a variety of issues that relate to results, impact domains and causal factors, as well as side effects. One of the first challenges is therefore to set up priorities in case studies between these competing purposes. Computing productivity at farm level is unlikely to be a priority. Conversely, any case study is likely to cover beneficiaries with/without productivity change over a given period, and interviews may be designed to allow a comparison of those two groups. Rather than looking at the completed first half of the RDP only, it may be useful at mid-term to cover prospects for change over the full RDP period.

Participatory approaches to rural analysis are based on the principle that the farm holding managers are critical decision-makers. These beneficiaries are therefore in the best position to explain to a well-meaning interviewer whether they have made investments or changed labour inputs or quantities produced, or plan to do so in the following years. Evidence can be semi-quantified as soon as samples are above 100 holdings and provided answers are partly collected in the form of scores (e.g., from 0, not at all, to 5 absolutely).

The capacity of case studies to confirm causal factors (or side effects) depends on (a) the appropriate definition of these factors, and (b) the sampling strategy. Additional indicators can be collected through the case studies. The 2000-2006 had highlighted the need to examine indebtedness, since over-indebtedness is an unwanted side effect of investment measures in mainland France. One of the critical factors in understanding the role of the RDP in 2007-2013 in productivity changes will be access to innovation. In addition to result indicator 3 in the common evaluation framework, a range of additional indicators deserve being tried:

- Indebtedness: annual reimbursements/farm income (over-indebtedness = unwanted side effect)
- Access to innovation: through/outside RDP; individual/collective; number of holdings/enterprises introducing new products and/or new techniques (Result indicator 3), introducing new management modes

4.3.4 Data requirements and collection

Data can origin from primary and secondary sources. The choice depends on the availability and quality of data as well as the cost-effectiveness if opting for primary data collection through surveys. Secondary data sources are the FADN, national or regional accounts, national farm accountancy networks. For agricultural holdings the test farm data (FADN) are a good source – with certain limitations: First they do not provide a constant data set as the test farms change every year. Second, they also do not include information on diversified activities of farms.

However, similar network data do not exist for the forestry or food sector. It is therefore required to rely on accounting data provided by beneficiaries, for example as part of reporting requirements, as well as on additional surveys. Business plans are only indicative for the expected business outcome; therefore the use and interpretation of this data should be made with caution. Accounting data from compulsory or test bookkeeping of beneficiaries and non-beneficiary holdings can be used; however experiences from the previous funding period have shown that data have been inaccurate and incomplete and

required additional efforts before being processed and interpreted. Possible difficulties might occur when collecting data from indirect beneficiaries. Once indirect beneficiaries have been identified, the purpose of data collection needs to be communicated in order to gain the good will of non-beneficiaries not underlying compulsory bookkeeping.

Real case example of assessment of programme impacts

Belgium: Impact Assessment of Rural Development Programmes: the Flemish experience.

Katrien Van Dingenen (IDEA Consult). A vertical approach: effect themes as a starting point for impact assessment.

For the ex-post evaluation of the Flemish RDP 2000-2006, the contracting authority explicitly asked the evaluation team to focus its efforts on analysing the **main impacts** at **programme level**. This focus and openness to prioritization allowed the evaluating team to adopt an impact-oriented approach. The evaluation team opted for a vertical approach, in which the various measures were analysed for their (partial) contribution to the **overall programme effects** regarding a number of specific core themes/objectives such as income, employment, soil erosion, water contamination, etc.

The possible effects are classified into **three thematic components**. Within these, 16 relevant themes were distinguished for evaluation:

- (i) agro(food) economy: income, employment, labour productivity and quality within agriculture, horticulture and the agro-food sector;
- (ii) the environment, nature and the landscape: soil, water, flora & fauna, landscape, ammoniac, greenhouse gases, water consumption and nuisance;
- (iii) rural quality of life: indirect employment, physical facilities, activities & services and perceived value of the physical environment.

The contribution from various relevant measures to each of these aspects was evaluated on the basis of the **proportionality principle**: the main research efforts were devoted to those measures which were expected to have a high **impact**. This "expected impact exercise" has been done on the basis of an interactive workshop, the available take-up information and existing evidence of causal relationships (from not necessarily RDP-related empirical research). The impact assessment exercise was based on **case studies** at **micro/local level** tracing net effects of the RDP's contribution to e.g. productivity.

4.3.5 Interpretation and judgement issues

4.3.5.1 The explanatory power of the labour productivity indicator

GVA/FTE is an area-sensitive, partial productivity measure which allows for intra-sectoral comparisons within a Member State. It is not suitable to express competitiveness of one sector against another one.

The indicator, if accumulated at EU level can express the competitiveness of the respective sector (at EU level) in comparison to other areas; the comparison and aggregation across Member States requires using Purchasing Power Standards as price basis.

The labour productivity indicator expressed in GVA/FTE is best used to answer the common evaluation questions on the improvement of the competitiveness of holdings in the agriculture, forestry or food sector. However, no feasible methods are in place yet to establish the contribution of a programme to the overall competitiveness of the relevant sectors.

Labour productivity can also be used to express the "economic performance of agricultural holdings" at constant prices; assuming that the increase in GVA at constant FTE or the reduction of FTE at constant

GVA leads to an increase in benefits for the holding, or that the maintenance of the ratio GVA/FTE also stands for maintaining of the level of economic performance.

The "better use of production factors in agricultural holdings" can either be a result of better use of physical capital assumedly increasing the GVA, or a result of an increase in labour productivity (increasing the amount produced per FTE).

4.3.5.2 Judgement issues

The indicator might be biased by drastic changes in price levels, entailing value changes of (produced) stocks. Such fluctuations should be smoothed out. An additional complication might occur through the fact that production methodologies and technologies used sometimes depend on the current price levels: *"A process which is economically efficient and profitable at one set of prices may cease to be so at another and would, therefore, not be used at those prices. For this reason, figures of gross value added obtained by revaluing the quantities at very different sets of relative prices may have little economic significance and may even become negative"* (UN 1993).

The indicator does not allow for taking side effects into consideration, for instance if funding is provided to companies whose improved performance makes little or no direct contribution to rural development in the localities from which they source their products or employees. The net effect would in this case be just a loss in terms of employment.

The indicator value does not allow a direct inference in respect to the EU Treaty objective of *"increasing agricultural productivity in order to ensure a fair standard of living for the agricultural community, in particular by increasing the individual earnings of persons engaged in agriculture"*. An increase in labour productivity does not allow the conclusion that the farmer has achieved a better living standard. Case studies and surveys may reveal the kind of interdependencies which are actually at work.

Resuming, the explanatory power of the indicator Labour Productivity has limitations because it consists of two factors: GVA and FTE, and thus it is not directly obvious which component actually induced the change in labour productivity. In order to overcome the limitations of the GVA/FTE indicator as measurement for labour productivity, the competitiveness of the agricultural sector can be measured in alternative ways, such as

- ▶ Competitive Performance: Frequently-used measures of competitive performance are market share (share of total sales from all sources in a given country or region, or of total exports of a particular product); or trade balances.
- ▶ Revealed Comparative Advantage (RCA): Value added of agricultural sector in the province/Gross domestic product of the province/Value added of agriculture sector of the whole country/Gross domestic product of the whole country;
- ▶ Growth Competitiveness Indicator (GCI) and Business Competitiveness Indicator (BCI): these represent causal approaches measuring factors influencing competitiveness such as the institutional environment, infrastructure, macroeconomic stability and cost structures;
- ▶ Domestic Resource Cost (DRC) using the Policy Analysis Matrix (PAM)⁷⁹, PAM being the product of two accounting parameters (profitability; and effects distorting policies and market

⁷⁹ Monke and Pearson (1989).

failures). DRC compares the cost of domestic resources to value added (both at social prices); it is based on typical average data for a sector (World Bank 2008). If DRC lies between 0 and 1, there is comparative advantage: the value of domestic resources used in production is less than the value of foreign exchange earned or saved.

Another possible definition for increased competitiveness can be the significance of agriculture/forestry in regional/national value chains. However, there is no straightforward indicator available related to this definition.

Alternative measures of labour productivity are multi-/total- factor productivity measure for performance comparisons of the agricultural industries between Member States using agricultural industry output at basic prices to a unit input bundle comprising capital, intermediate consumption and labour.

5. The assessment of environmental RDP impacts

In comparison to the assessment of socio-economic impacts, assessing the impact of RDPs on the environment poses a number of intrinsic methodological challenges, among which:

- ▶ The sub-sequential effects of rural development measures: firstly on the behaviour and management practices of farmers and forest holders, and, secondly, in terms of impact on the environment due to the changed farming/forestry practices;
- ▶ Impacts are often depending on site-specific circumstances, such as soil, temperature, rainfall etc. As a consequence, linking the results of on-site observations to overall conclusions at the level of the programme area is not a straightforward task;
- ▶ Impacts may take a long time to emerge. Therefore the assessment should preferably make use of long-time series data, where these are available;
- ▶ Due to complexity and site specific impacts of RDPs on the environment, the identification of control groups and the establishment of a situation with and without the programme in place (counterfactual situation) are particularly difficult;
- ▶ In the context of rural development programmes, a broad range of measures, from different axes, may affect the environmental conditions of a given programme area;
- ▶ It is often difficult to establish cause-effects relationships for environmental impacts.

The sections which follow intend to provide hands-on advice on how to tackle these challenges in the context of the evaluation of the rural development programmes. Each of them deals with one of the specific themes addressed by the four common impact indicators defined by the CMEF, respectively.

The following can be considered as general recommendations for facilitating the assessment:

- ▶ The establishment of programme-specific additional indicators is a key element for ensuring a correct and comprehensive assessment of environmental impacts;
- ▶ Appropriate monitoring and ongoing evaluation activities are essential for ensuring that the evaluators dispose of relevant data on general trends, outputs and results;
- ▶ Exploiting synergies and complementarities with alternative sources of information can improve the cost-effectiveness of data collection activities, as well as the quantity and quality of the data available for the evaluators;
- ▶ The collection of relevant primary data by the independent evaluators, and their skills in complementing the quantitative data available with qualitative information and sound expert judgments, represent central ingredients for a well founded estimation and quantification of environmental impacts.

5.1 Impact Indicator 4: Reversing Biodiversity Decline

5.1.1 The CMEF guidelines

The following sections are supposed to set the frame for depicting effects of RD programmes on “reversing biodiversity decline”. The assessment of impact in this field is composed of combining information from baseline, result and impact indicators.

5.1.1.1 Common baseline indicator no 17: Population of farmland birds

The farmland bird indicator (FBI) is intended as a barometer of change for the biodiversity of agricultural landscapes in Europe. Assuming a close link between the selected bird species and the farmland habitat, a negative trend signals that the farm environment is becoming less favourable to birds.

FBI is an aggregated index of population trend estimates of a selected group of 19 breeding bird species dependent on agricultural land for nesting or feeding. The following farmland bird species are included: *Alauda arvensis*, *Burhinus oedicephalus*, *Carduelis carduelis*, *Columba palumbus*, *Emberiza citrinella*, *Falco tinnunculus*, *Galerida cristata*, *Hirundo rustica*, *Lanius collurio*, *Lanius senator*, *Limosa limosa*, *Miliaria calandra*, *Motacilla flava*, *Passer montanus*, *Saxicola rubetra*, *Streptopelia turtur*, *Sturnus vulgaris*, *Sylvia communis*, *Vanellus vanellus*.

However, MS may use an alternative composition of bird species where this is appropriate to national/regional situation. This alternative is important because the set of species used in the European Bird Census Council (EBCC) Pan-European indicator is not entirely applicable in all MS.

Data on individual species population is collected annually through surveys. The population counts are carried out by a network of ornithologists (mostly volunteer) coordinated within national schemes.

Indices are calculated for each species independently and are weighted equally when combined in the aggregate index using a geometric mean. Aggregated EU indices are calculated using population-dependent weighting factors for each country and species.

The indices are compiled by Statistics Netherlands in conjunction with the Pan-European Common Bird Monitoring scheme (PECBM: a joint project of the European Bird Census Council, the Royal Society for the Protection of Birds, BirdLife International, and Statistics Netherlands).

For the purpose of comparability the Commission has chosen a common reference year where maximum geographical coverage is provided. FBI is indexed on the year 2000, However, Member States may choose other years where this improves the quality of the analysis.

5.1.1.2 Result indicator – “area under successful land management contributing to biodiversity and high nature value farming/forestry”.

According to Guidance Note I of the CMEF the result indicator is “area under successful land management contributing to: (a) biodiversity and high nature value farming/forestry”. The indicator measures the total amount of hectares under successful land management. Successful land management is defined as the completion of land management actions contributing to:

- ▶ Improvement of biodiversity
 - protection of wildlife species or groups of species
 - protection or restoration of natural and semi-natural habitats
 - maintain reintroduce crop-combinations
 - safeguarding endangered animal breeds and plant varieties

It is crucial that there is a real link between biodiversity and ‘successful’ land management. The MS should be able to show which agri-environmental measures or types of land use (both cropped and uncropped) are really contributing to the improvement of biodiversity and to which degree. Additional result indicators should be formulated accordingly after having implemented specific studies, as the way impact monitoring is organised in the UK-England.

Good practice for assessing impacts on measure level by combining a multitude of data, like those of FBI, ongoing monitoring and special studies an example from UK:

The UK smoothed farmland bird index of 19 species would appear to provide the information necessary to report against reversing the long-term decline in farmland bird populations. However the problem with just using this information to measure progress against the Impact Indicator is that the populations of these species are determined by many factors. There has been a 48% decline in the index since 1970. It is generally accepted that multi-faceted agricultural intensification has been the major driver in the long-term decline in farmland bird populations, although the individual factors differed between species and geographically. The index appeared to level off in the late 1990s but has since begun to fall again coincident with widespread AES provision. The reasons for this recent decline are unclear but could relate to a range of farming (e.g. the ending of compulsory set aside in 2007, decline in bare fallow land) and non-farming (e.g. climate/weather affects, increase in predators) related factors. Against this background it is not easy to identify the contribution being made by the Rural Development Programme Measures – the recent declines might have been far more severe in the absence of AES provision. To isolate the effects of the measures from other contextual factors evaluators’ expertise is essential. The interpretation could be easier if the monitoring program include pair-wise comparisons with control sites.

It has to be accepted that it is not yet possible to build up a fully comprehensive and scientifically rigorous picture of the impact of the English Rural Development Programme on farmland birds. We do however have a number of the components of this picture. These are as follows:

1. The results of autecological studies showing a positive response to targeted agri-environmental management by rare and localised species such as stone curlew, curlew and black grouse.
2. The farmland bird index and the national population data for each of the 19 species that make up the index.
3. A wealth of scientific evidence that can be used to construct a ‘change of causality’ to link output information (the provision of suitable habitat), result information (demonstration of benefit at the option or farm scale) and species population data to provide an estimate of the contribution agri-environment schemes are making to sustaining the populations of the more widespread farmland bird species.

It is suggested that progress towards achieving this indicator might be measured using three parameters:

1. The farmland bird index
2. A measurement of the area of habitat being provided under Axis 2 RDP Measures that is known to be of value to the farmland bird species tracked in the index
3. Direct measurements of the populations of rare and localised farmland bird species known to have benefitted from agri-environmental management.

When reporting these three measurements it would also be necessary to include a short commentary, in particular to explain some of the other factors that may have affected farmland bird populations during the Programme period.

National populations of many widespread farmland birds are already monitored, but it has always proved difficult in the past to relate the observed trends to agri-environmental management. However, by 2013, there will be population data covering more than 7 years of large-scale agri-environment management in England. With this and the data collected from the previous links in the chain, it should be possible to analyse the population data for individual species and to correlate observed population changes against changes in the deployment of agri-environmental management designed to benefit these species. This analysis should then provide, in combination with the data for the rare and localised species, the best possible measurement of the extent to which RDPE has contributed towards the ultimate target of reducing the decline in farmland bird populations.

Data source: (Geoff Radley, Evidence Team, Natural England)

5.1.1.3 Impact indicator – change in trend in biodiversity decline as measured by farmland bird species population

Guidance Note J of the CMEF defines this impact indicator as “change in trend in biodiversity decline as measured by farmland bird species population”. It is then described as “change in trend in biodiversity decline in the area targeted by the intervention as the quantitative and qualitative change in species population that can be attributed to the intervention once double counting, deadweight, and displacement effects have been taken into account.

Farmland bird species population is an indicator of general biodiversity trends for which the best data exist in term of time series and geographic distribution. It can be complemented by other existing indicators such as population trends of agriculture related butterfly species, or trends in important bird areas (IBAs) considered as threatened by agricultural intensification, and under-utilisation of land or abandonment.

Member States may wish to make use of other national or regional indicators to further interpret changes in the population of particular bird species characteristic of the programme area. MS may use an alternative composition of bird species where this is appropriate to national/regional situation.”

Although this paper focuses on breeding birds it is important to remember that if there is an intervention in the frame of agri-environmental measures which is supposed to have positive effect on non-breeding, wintering or migrating birds (e.g. geese, swans, waders, seed-eaters etc.), such bird species should be targeted in monitoring and evaluation.

5.1.1.4 Measures and evaluation questions related to these baseline, result and impact indicators

The community strategic guideline states that: “To protect and enhance the EU's natural resources and landscapes in rural areas, the resources devoted to axis 2 should contribute to three EU-level priority areas: biodiversity and the preservation and development of high nature value farming and forestry systems and traditional agricultural landscapes; water; and climate change”.

The measures available under axis 2 should be used to integrate these environmental objectives and contribute to the implementation of the agricultural and forestry Natura 2000 network, to the Göteborg commitment to reverse biodiversity decline by 2010, to the objectives laid down in Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (1), and to the Kyoto Protocol targets for climate change mitigation. (Council Decision 2006/144/EC, OJL 55/29).

Guidance Note B of the CMEF presents the common evaluation questions including horizontal questions. Horizontal questions to be answered in the context of the baseline indicator no 17 are:

- ▶ To what extent has the programme contributed to promoting sustainable development in rural areas? In particular, to what extent has the programme contributed to **biodiversity** for protecting and enhancing natural resources and landscapes in rural areas?
- ▶ To what extent has the programme integrated environmental objectives and contributed to the realisation of Community priorities in relation to the Göteborg commitment to reverse biodiversity decline?

Common evaluation question directly connected with the impact indicator is the question for measure 214 Agri-environment payments:

- ▶ To what extent have agri-environmental measures contributed to maintaining or improving habitats and biodiversity?

5.1.2 Key challenges with regard to measurement and interpretation

- ▶ **Data availability** – MS should find out data availability on birds for baseline (Farmland Bird Index, FBI) and impact indicators as early as possible. That includes investigation of existing projects coordinated or executed by the responsible ministries, research institutes, universities, environmental NGOs (in particular BirdLife International partners) etc.
- ▶ **Data reliability** – available data may in many cases not be reliable for assessing policies, so, the background of available data has to be documented properly (e.g. geographical coverage, sampling strategy, the number of replications, data collection methodology etc).
- ▶ **Data representativeness** – the representativeness of available data should also be investigated (e.g. statistically by using specific analysis, expert opinion asked, etc), especially considering farmland under agri-environmental schemes (AES).
- ▶ **Consideration of regional differences** – while using reliable and representative data on birds in evaluation, regional differences (within and between Member States) should be kept in mind. List of species should be modified as to represent biodiversity at regional or local scale. While doing that the general checklist of 116 farmland bird species should be followed to avoid non-farmland species within their FBI.
- ▶ **Counterfactual situation** – A big challenge will be to find an appropriate comparison group (as there are no identical situations in the nature). So, multiple factors which may affect the outcome have to be taken into account which is quite complicated, especially in the case of 'broad and shallow' measures affecting a large proportion of land area. GIS based bird data collection and spatial modelling approach for assessment of several/many explanatory factors simultaneously should be used where possible and always encouraged. In some MS, e.g. Austria, Finland and Sweden the formation of comparison group for agri-environmental measures is practically impossible as they cover 90 and even more per cents of agricultural land.
- ▶ **Net effect** – To separate the net effect of support from the gross effect, it is important to analyse exogenous factors effecting beneficiaries, but not depending on the RDP implementation. So, additional data (details of site location, crops, managements, habitats and

landscape etc) are needed. In addition, usually, multivariate statistics are needed to assess the net effects of RD measures which is a challenging task.

- ▶ **Double counting, deadweight and displacement** – Double counting is the situation where the objects are counted more than once – as the methods are state of the art in bird science (if regarded to be constant) double counting is not a big problem. Deadweight and displacement effects may be difficult to quantify and may, at best, be addressed in a qualitative and contextual manner or with demanding multivariate approaches.

5.1.3 Recommendable methods of measurement

5.1.3.1 Common baseline indicator no 17: Population of farmland birds

A measurement of farmland bird species population trends is made by the Farmland Bird Index (FBI). FBI has been adopted as an EU Structural Indicator and a Sustainable Development Indicator. The FBI is intended as a barometer of environmental change in the countryside and thus as a surrogate for changes in biodiversity and nature more broadly. It is designed to be sensitive to a number of different potential drivers and pressures on the environment.

FBI is a multi-species index obtained by the aggregation of a set of individual species indices using a geometric mean. Individual indices are calculated for each species independently. By using the geometric mean, the species are weighted equally in the indicators. In case, the species indices are provided for a time period of a different length, the chain method is used in the indicator computation.

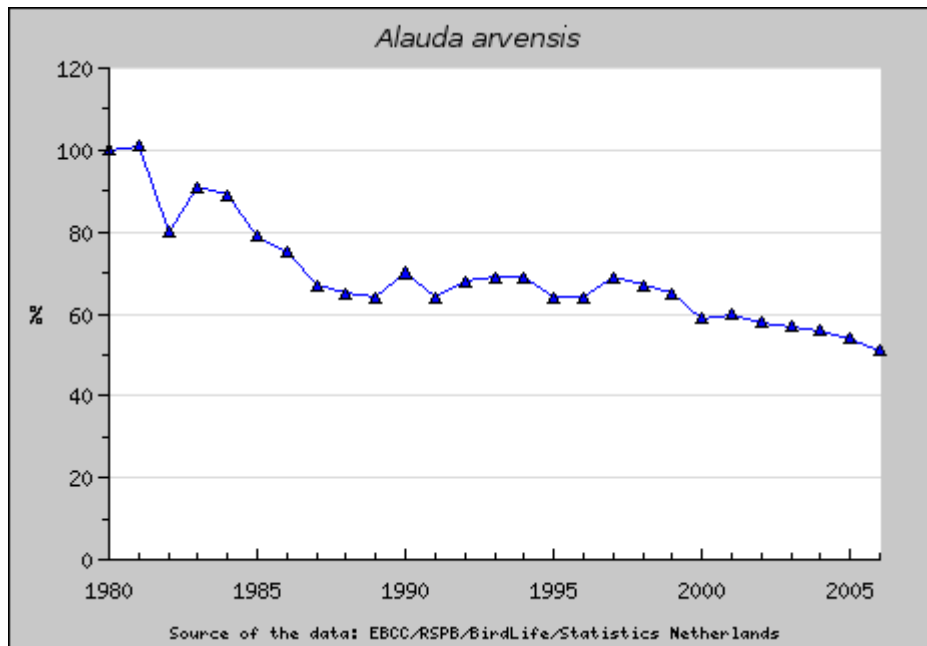
The population trend data are based on a selection of 36 bird species breeding in and characteristic of farmland habitats in Europe. In 2005 the list of species covered was modified to be more specific to farmland: 19 breeding bird species dependent on agricultural land for nesting or feeding were selected.

Weighting allows for the fact that different countries hold different proportions of each species' European population.

EU FBI is indexed on the year 2000, this base year having been selected so as to provide the maximum geographic coverage. The change is measured backwards and forwards from this base year

The data are presented as population trends of indices both of separate species and of multi-species, as shown in Figure 5. The trends result of statistical modelling to allow estimations of numerous sites with locally discontinuous data series

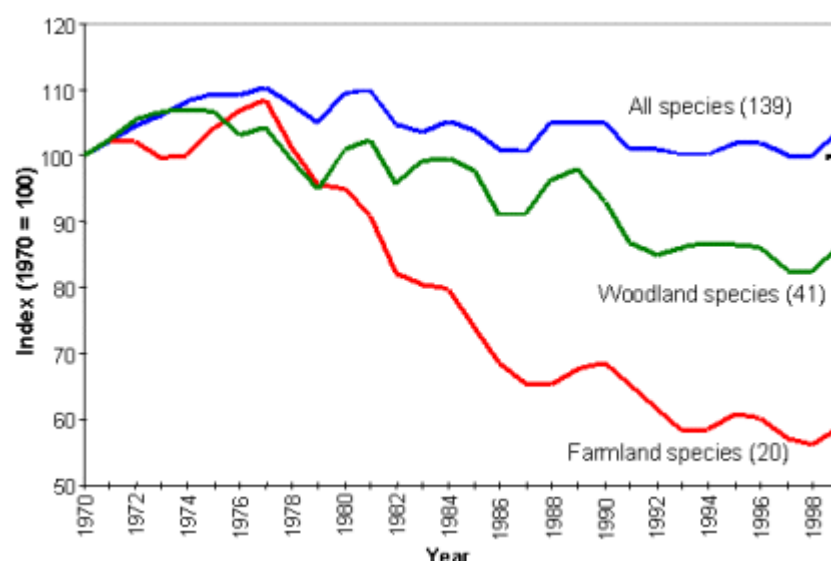
Figure 5 An example of single species index trend in Europe. *Alauda arvensis* (Skylark) is a typical farmland bird species.



Data source: <http://www.ebcc.info/index.php?ID=360&species%5B9760%5D=1>

Aggregated European indices are calculated using population-dependent weighting factors for each country and species (estimates of national population sizes (derived from Birds in Europe 2, 2004 see Figure 6).

Figure 6 An example of aggregated indices trends



Data source: <http://www.ebcc.info/index.php?ID=28>

Based on their local conditions, MS and regions may need to use an alternative composition of bird species where this is appropriate and may also to choose a different reference year where this improves the quality of the analysis, or changed if it becomes clear that the year 2000 in the MS is in fact an extraordinary year; as is suggested in the following box.

MS may use their own national baseline indicator – an example from Finland

Finland has decided to use biodiversity baseline indicator no 17B “Bird indicator based on the ecological grouping of birdlife nesting in farmland”. The indicator is defined as the average index of about 40 species and can be ecologically subdivided into species feeding in farmland and breeding in arable areas, field margins, forest areas or farmyards. Ecological grouping helps to identify the impacts in greater detail because species in different groups experience farming work, management and land use differently.

Data sources: RDP for Mainland Finland 2007-2013,
http://www.mmm.fi/en/index/frontpage/rural_areas/ruraldevelopmentprogrammes/strategyandprogramme20072013.html

Managing the index according to availability of new data and methodologies and eventually sub-dividing it into sub-indices may also have additional value for the interpretation of changes, as the following example from DEFRA-UK suggests.

FBI – management and subdivision of index – an example from UK-England

Defra has a UK Government target to reverse the long-term decline in the farmland bird populations by 2020. Progress against this target is measured annually using a smoothed index of the populations of 19 widespread breeding bird species that are closely associated with English lowland farmland. Smoothed species indices are used in order to reduce the level of noise in the indicator and reveal underlying trends.

An updated farmland bird indicator for England has been produced for the period 1966 to 2005, using joint CBC/BBS (<http://www.bto.org/survey/complete/cbc.htm>) or anchored trends as appropriate for 19 widespread breeding bird species within the indicator. The index is derived by modelling and estimates are revised when new data or improved methodologies are developed and applied retrospectively to earlier years.

Farmland birds index comprises 19 species of common and widespread birds closely associated with farmland habitats out of which *Motacilla flava*, *Sylvia communis*, *Vanellus vanellus*, *Falco tinnunculus*, *Carduelis carduelis*, *Alauda arvensis*, *Columba palumbus*, *Passer montanus*, *Streptopelia turtur*, *Emberiza citrinella*, *Sturnus vulgaris* are common with European FBI species and *Carduelis cannabina*, *Corvus frugilegus*, *Perdix perdix*, *Carduelis chloris*, *Emberiza*

calandra (*Miliaria calandra*), *Corvus monedula* and *Columba oenas* are specific.

The index can be subdivided into separate indices for generalist and specialist species (those that breed mainly or solely on woodland/farmland), species are identified for the woodland and farmland indices.

Data sources: http://www.bto.org/research/indicators/farmland_psa.htm,
<http://www.defra.gov.uk/evidence/statistics/environment/wildlife/wdbirdspop.htm>

5.1.3.2 Impact indicator – change in trend in biodiversity decline as measured by farmland bird species population

In principle the impact indicator 'change in trend in biodiversity decline' as measured by farmland bird species population is measured in a bottom up way – linking the single measure of the RD programme with the effect on the farmland bird species population thus establishing a cause-effect relation. For that purpose it is essential to find out as early as possible which data on bird impact indicators are available and which are the specific studies to be carried out.

In certain occasions capturing impacts at measure level can be done by using FBI data on a national scale, although they are not intended to measure the effectiveness of specific, fine-scale measures (e.g. agri-environment measure) implemented at site level.

If the measure in question is being implemented at a very broad scale (e.g. reducing pesticide inputs across entire farms), and there are enough sample plots in areas where the measure is applied it might be possible to use also the results of common farmland bird monitoring to assess the effectiveness of the measure. So, countries whose FBI data have a very good coverage may use these data also for the impact assessment of measures, as suggested by the example in Austria.

FBI data used for impact assessment on the measure level – an example from Austria

Austria has decided to use as much existing FBI information as possible also in capturing impacts at measure level. This would be possible because in Austria AE measures have a coverage of app. 90% of agricultural land. Additional sampling has been set up (both professional counters and volunteers) to achieve big enough sample sizes for all indicator species. The FBI will be subdivided after several aspects (farming type arable/grassland, Natura 2000, LFA, etc) to allow deeper insights – as long as possible due to the sample sizes of the indicator species. This calculation will be done in 2010 for the mid-term evaluation.

Data sources: BirdLife Austria, Norbert Teufelbauer

In most cases the FBI does not however have such a good coverage or the data do not coincide with areas under specific measures. In these cases the existence and use of other previous or ongoing bird monitoring should be investigated. In addition, there may be the need to carry out special studies to get data in order to evaluate the impact of measures. There is also a possibility to combine the data of existing common bird monitoring and special measure specific studies.

The FBI indicator can provide an overall picture of the impact of the RDP. In most cases, the evaluation of individual measures or schemes requires ad hoc and highly replicated field studies, including pair-wise comparisons with control sites. That is a comparison between areas that differ mainly for the application or not of the measure. It is very important, however, to 'control' other influences, which is possible by careful design of sampling and use of multivariate analysis in most cases. It is also necessary to partition environmental variability that is not directly linked to the measure, making it possible to separate effects actually caused by the measure on birds' populations from other spurious effects. This is the best way to

demonstrate with a high degree of certainty that a particular measure is being effective, and why. However, it should be assessed (e. g. by careful study design) to which degree such studies are representative for the whole area where the measures under investigation are implemented.

The choice of species/communities to monitor should be governed primarily by the specific objectives of the individual measure, which in turn should always consider how they could inform, and impact EU and national indicators and priorities.

Evaluations should take place at the field scale (to assess the benefit of individual measures) and/or at landscape scale (to assess wider-scale populations benefits), as appropriate for the species being studied. For species that disperse widely (for example most birds and bats) both scales may be of interest. For species that are less dispersive and slow to reproduce, it is essential that they are studied at the field-scale or patch-scale, as populations will respond directly to localized land management.

Survey and analytical techniques will be highly variable between individual species/communities and may vary with other factors such as topography and time of year. Expert opinion should always be sought when designing monitoring strategies.

The following three examples from various MS (EE, AT, FI & UK-ENG) shall highlight the need for highly specific approaches when applying the common impact indicator in the different settings:

What to do when no data is available – which data would help to evaluate the impact of the measures what kind of special studies should be carried out – an example from Estonia

Monitoring of populations of farmland birds in Estonia has challenges in reflecting general status of farmland biodiversity (baseline situation) and to fit to RDP or measure specific evaluation – for the calculation of FBI only few counting areas are used which also have small coverage of agricultural landscapes.

For capturing the impacts of Estonian AE measure of RDP 2004–2006 period a special farmland bird monitoring (taking the number and density of breeding birds into account) was started in 2005 in the frame of agri-environmental evaluation in Estonia. Together with other biodiversity data (bumblebees, vascular plants and earthworms) farmland birds are monitored also for the RDP 2007–2013 period in 66 farms in total covering different regions (reflecting different soil and weather conditions and agricultural intensities), farm sizes and farm practices (e.g. organic farms and conventional farms). Monitoring sample also consists of reference farms not participating in the agri-environmental scheme. Depending on the region and landscape structure the composition of bird species may vary quite significantly in Estonia. Birds are counted annually using the line transect method (0.5-2 km per field). Farmland bird data are analysed together with data for other taxa (especially bumblebees) and landscape.

In addition, the relationship between agricultural production and the possible impact of the agricultural support system on the bird population is studied in one of the pilot areas located in Räpu catchment.

Data source: Agricultural Research Centre, <http://pmk.agri.ee/pkt>

Possibility to combine data of existing common bird monitoring and special measure specific studies – an example from Austria

For RDP 2000-2006 period Austria used special field study to evaluate agro-environmental effects on biodiversity (wildlife species, habitat diversity and landscape). Existing sampling data (e.g. specific survey conducted in 1998) were used as much as possible to show the impact of the measure over a specific period of time. Hence, 10 sites covering two main Austrian landscape types with an area of 1 km² each were sampled repeatedly. At these spots birds, plants and landscape features were mapped again with the consistent method in 2003. Similarly, the spatial allocation of agri-environment measures in these sampling plots was documented. Thus an analysis of the impact of the agri-environment measures was possible as a time-based approach as well as a location-based approach: The development of biodiversity elements over a time period of five years was correlated with the applied agri-environment measures and the effect on biodiversity and landscape features was quantified. Additionally, a simultaneous comparison of the biodiversity status between areas with and without specific agri-environment measures was carried out.

This sophisticated approach responds to species diversity as well as to habitat and landscape diversity. The statistical analyses of the results provide sound information about the impact of the measures. It was concluded that for species diversity and habitat diversity field surveys were the best method. Enlargements of the sampling sites & increased numbers of samples are necessary and will certainly enhance the validity of the results.

This study (comparison 1998-2003) did not distinguish between the effect that some areas are a priori less intensively used and the effects of the agro-environmental measures itself. There was not multivariate approach behind, and certain 'control variables' (as soil productivity and other) were not taken into account. Therefore, the result that areas with higher percentage of certain agro-environmental measures ('light green' ones) are associated, e. g., with higher bird densities, is not conclusive, as agro-environmental net effects were not identified; these were simply less productive areas, where the uptake of 'light green' and other measures is higher, because in practice they do not restrict farming very much. This 'mistake' is a very common one in evaluation studies.

Data source: E.Schweiger's presentation <http://pmk.agri.ee/pkt/CD/index.php?page=2>, Comments of Johannes Frühauf – BirdLife Austria

Agri-environmental monitoring studies may produce positive side effects by offering baseline data and creating base for a comprehensive farmland biodiversity monitoring – an example from Finland

The efficiency of Finnish agri-environment support scheme has been studied in the MYTVAS 2 research project (2000–2006). Nature-Mytvas aims to estimate the effects of the supported agri-environmental measures on farmland biodiversity and landscape. It also produces baseline data on the level of biodiversity in ordinary Finnish agricultural areas on several taxa – birds, vascular plants, butterflies, day-active moths and bees. This creates a solid base for long-term monitoring of Finnish farmland biodiversity. Nature-Mytvas is further divided into two major parts: a large-scale species monitoring project conducted on a large group of randomly selected study sites, and several smaller case studies on the biodiversity effects of specific supported measures. MYTVAS3 continues the efforts during 2008–2013.

Data source: <http://www.environment.fi/default.asp?contentid=200556&lan=fi&clan=en>

5.1.4 Data requirements and collection

5.1.4.1 Common baseline indicator no 17: Population of farmland birds

At present the data for the calculation of FBI originate from national monitoring of widespread birds collected and compiled by the Pan-European Common Bird Monitoring Scheme (PECBMS) in cooperation with Statistics Netherlands. In 2008 21 European countries were involved directly in the collection of data for the index: Austria, Belgium, Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Latvia, Netherlands, Norway, Poland, Portugal, Spain, Sweden, Switzerland and United Kingdom. Information on national monitoring schemes, indicators and the use of their results can be found on the EBCC web site (www.ebcc.info).

Most EU countries have well established and high quality common bird monitoring schemes running – by relying and building on these existing schemes, Member States can greatly reduce the cost and administrative effort of collecting, processing and interpreting bird data for the purpose of Rural development monitoring and evaluation. **Although national schemes differ in count methods in the field, these differences do not influence the supranational results because the indices are standardised before being combined.**

Two bird monitoring scheme examples, different in size:

MS that are quite at the beginning with the FBI data collection – an example from Slovenia

In Slovenia pilot scheme called „Slovenian monitoring of common birds of agricultural landscape“ exists since 2007. 15–25 fieldworkers conduct line transect counts (stratified systematic, habitat data available). Data are available for 21 bird species.

Data source: <http://www.ebcc.info/pecbm-slovenia.html>

MS with long histories with FBI data collection – an example from Germany

If a change in the basics of the monitoring programme is necessary, it is ideal to run the old and the new programme side by side for a period for assessment of comparability and calibration.

In Germany currently there are two relevant monitoring schemes. Former scheme (started in 1989) will be finished in 2010 and afterwards, the later scheme (started in 2004) will completely take over common breeding bird monitoring in Germany in the next years.

Old scheme: “DDA monitoring programme for common breeding birds”, there are 500 fieldworkers and data for 100–130 bird species are available. They use three different methods for data-collection: point counts, line transects and territory mapping. Selection of plots is mostly free choice; habitat data are recorded.

New scheme: “DDA Monitoring programme of common breeding birds in the wider countryside”, since 2004. There are 1300 fieldworkers and data for 100–150 bird species are available. Line-transect mapping is used (like the Swiss field method), the selection of plots is stratified random and habitat data are recorded.

Data source: <http://www.ebcc.info/pecbm-germany.html>

Data from generic monitoring schemes may often not be ideal for specific evaluation purposes, either because of poor geographical coverage of an area, or because they do not provide highly detailed site specific information on numbers and distribution required for these purposes. Therefore it is vital for the evaluator in the first instance to ask local monitoring data providers to verify the representativeness of their data, especially from farmland under agri-environmental schemes (AES) point of view. If it is clear that monitoring data do not represent the target area it should be clearly mentioned in the report and linking AES with FBI should be made very cautiously.

Sustainability of data collection – though the actual common birds' surveys are carried out mostly by volunteers and thus represent a unique low cost resource of data, some funding is still needed for coordination and support to the volunteers' network and for the data processing phase. **The reliability of the data** – it is important to have a large and representative set of sample points and species – in the case of an indicator based on multi-species index of changes in abundance, in general the more species contributing to the indicator, the more reliable it is.

On the other hand only good indicator species should be included (e.g. farmland specialists likely to be affected by changes to habitat quality, species easily detected in common bird surveys), remembering, however, that some species difficult to find may be the most vulnerable specialists).

Individually, many species may show annual changes in abundance that may reflect a variety of environmental factors, such as extreme weather conditions during the breeding season, poor conditions on the winter grounds, changes in predation pressure, and simple sampling error and statistical noise. Long term monitoring and use of a wider range of species helps detecting the underlying trends. It is the evaluators' expertise that is essential for isolating the effects of the measures from other contextual factors.

To guarantee quality of the national FBI it is necessary to have an appropriate national monitoring scheme. Often national monitoring schemes do not cover representative amounts of farmland and therefore are weak in assessing real changes in populations of farmland birds. FBI is presented as one number without explanations. So, it is not known what is behind the values submitted by the countries and what aspects should be taken into account in the case of each of them. E.g. in the case of Estonia a lot of FBI monitoring points are located in forests and not on agricultural land. Such problems could be overcome by the species selection – indicator species should show a strong linkage to farmed land. In addition, many FBI monitoring points are located in protected areas. Even the proximity of protected areas to agri-environmental schemes bird impact indicator monitoring places can affect the results: protected areas are favouring high number and species richness of birds which may disperse to agricultural land in the vicinity. In such agricultural land the results of bird population are overestimated. Such a problem occurs especially in the case of species which have large territories, e.g. Eurasian Curlew (*Numenius arquata*). Such problems can possibly be overcome in special studies with spatial modelling where characteristics of surroundings of the study plots are included in the set of explanatory variables.

The possibility to use an alternative composition of bird species where this is appropriate to national/regional situation – while FBI calculates the aggregated Europe wide indicator on a basis of a “basket” of species tailored to best capture Europe wide trends, its use at national or regional level requires “species baskets” tailored to the local conditions (i.e. including species that are good indicators for farmland habitats and are common enough to be captured in common birds surveys). So MS may use an alternative composition of bird species (can of course include more or less species than the European wide one) where this is appropriate to national/regional situation.

The following three examples from AT, IT & LV illustrate how MS use an alternative composition of bird species where this is appropriate to the national/regional situation. Meanwhile the FBI (which one) is very sensitive of the individual species included in the index, indices with different species composition indicate different trends. So, it is crucial to state which indicator is used for which purposes.

Alternative and appropriate composition of bird species for depicting the baseline situation

Austria is using their own species selection for FBI, taking into account species specialized on farmland in Austria and also farmland at high altitudes. They have also decided to use the whole time period available (not only timeline since 2000) to facilitate evaluation (the longer the time series, the easier it is to see what is happening). Austria has calculated FBI 1998–2008 with the data available, i.e. 20 out of 24 indicator species and few data on high altitude farmland. A more or less 'complete' FBI can be expected from 2008 onwards. Additional sampling has been set up (both professional counters and volunteers) to achieve big enough sampling sizes for all indicator species. For the midterm evaluation of 2007–2013 Austrian RDP the FBI will be subdivided into several classes to allow deeper insights – due to the sample sizes of the indicator species. The classes will be, e. g.:

- farming type – arable/grassland
- Natura 2000 area
- less favoured areas – LFA
- entire Austrian territory

Data sources: Birdlife Austria, Norbert Teufelbauer

Italy is using the FBI, but with a different series of species with the objective of having an index more representative

of biodiversity at the regional scale. In particular, the Ministry of Agriculture in cooperation with NGOs is defining regional lists of species in order to draw appropriate regional programmes. This because Italian farmland is very different among regions in landscape, natural elements, agricultural practices, climate, etc; and bird species & populations vary widely among regions.

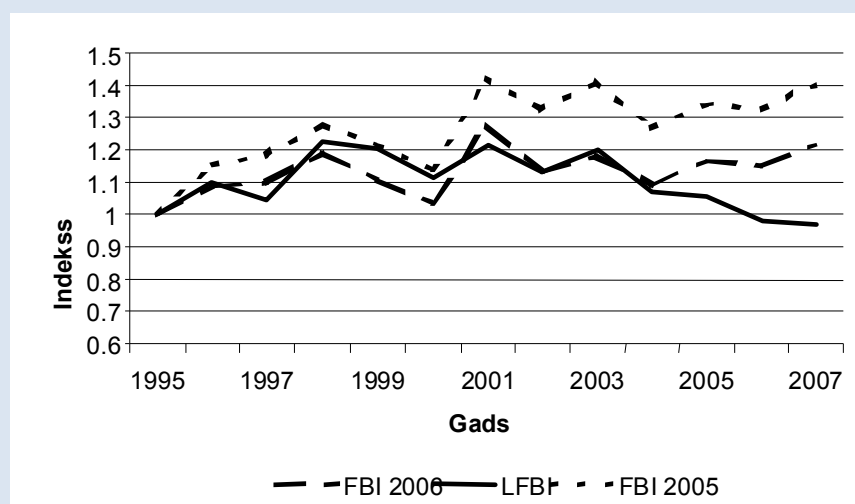
Using existing data the existing basket of species at national level is being revised to create a new basket of farmland species appropriate at regional level. On the basis of the new species basket and on the land use data related to census points (these data are available because the bird monitoring is in place since 2000), only points characterized by farmland habitats have been included in the new census program. In this way, it is almost sure that species in the index are related to farmland habitats.

Medium environmental characteristics of points where a species is detected is calculated for all species, than species are clustered. Species in each cluster show similar environmental parameters. Statistical analyses are then run in order to find the habitats closer to each cluster (principal components analysis, reciprocal averaging, and non-metric multi-dimensional scaling). Species for FBI are those in the clusters closer to agricultural habitats. Results of statistical analysis are validated with an expert based approach. This analysis can be done because common bird monitoring is carried out since 2000 and because habitat data are collected together with bird data.

Data source: Italian League for the Protection of Birds (LIPU), partner of BirdLife International, Patrizia Rossi

Latvia has faced the issue that some species used for calculating the FBI occur locally in other habitats and that may give different reflection also on FBI. In Latvia FBI has been calculated since 1995, using data on Farmland Bird Monitoring. Three versions of the Farmland Bird Index are calculated (fig. below): (1) Latvian Farmland Bird Index – calculated by combining bird species important for farmland habitats; (2) European Farmland Bird Index in Latvia – calculated with EBCC species list for Farmland Bird Index 2005 and (3) European Farmland Bird Index in Latvia – calculated with EBCC species list for Farmland Bird Index 2006.

Three versions of Farmland Bird Index (FBI) in Latvia⁸⁰:



The indices are different on the basis of species included. In Latvian Farmland Bird Index species like Barn Swallow (*Hirundo rustica*), Starling (*Sturnus vulgaris*) and Tree Sparrow (*Passer montanus*) were excluded, since they are mostly breeding at the human settlements but not on farmland in Latvia. Common Whitethroat (*Sylvia communis*) and Red-backed Shrike (*Lanius collurio*) were excluded because both species are mostly breeding in bushes and young forests in Latvia. However important species in Latvian farmland – Grasshopper Warbler (*Locustella naevia*), Marsh Warbler (*Acrocephalus palustris*), Goldfinch (*Carduelis carduelis*) and Common Rosefinch (*Carpodacus erythrinus*) were included

⁸⁰ LFBt – Latvian FBI, including White Stork, Corncrake, Lapwing, Skylark, Meadow Pipit, Grasshopper Warbler, Marsh Warbler, Whinchat, Goldfinch, Linnet, Common Rosefinch and Yellowhammer; FBI 2005 – FBI in Latvia according to EBCC 2005 list: White Stork, Corncrake, Lapwing, Skylark, Barn Swallow, Grasshopper Warbler, Marsh Warbler, Common Whitethroat, Whinchat, Red-backed Shrike, Starling, Tree Sparrow, Linnet and Yellowhammer; FBI 2006 – FBI in Latvia according to EBCC 2006 list: White Stork, Corncrake, Lapwing, Skylark, Meadow Pipit, Barn Swallow, Common Whitethroat, Whinchat, Red-backed Shrike, Starling, Tree Sparrow, Linnet and Yellowhammer.

in Latvian Farmland Bird Index.

Data source: O. Keiřs' presentation <http://pmk.agri.ee/pkt/CD/index.php?page=2>

5.1.4.2 Impact indicator – change in trend in biodiversity decline as measured by farmland bird species population

If monitoring aims to evaluate changes over time (either within measure sites, or relative differences between measure and control sites), then evaluations must take place over an appropriate length of time (dependent on species/community being evaluated). Such evaluation may be of the greatest value when comparisons can be made against 'baseline data' (data collected prior to the implementation of the measure).

The collection of data should enable the comparison between sites where the measure is being implemented and controls (otherwise similar sites where the measure is not being implemented). If the species/community to be monitored is known to be scarce or range restricted, then selecting survey sites entirely at random may be undesirable, as many survey units will be outside the species' range. In order to sample more effectively, stratification can be used to target areas within the known range. Survey sites (measure and controls) are then picked at random within delimited areas (or set percentages of survey sites are selected to fall within and out with the delimited areas).

Details of site location, crops, managements, habitats, landscape features etc within and around the survey units should be collected for use as co-variables in the data analysis, to account for variation not directly explained by the measure.

However the following challenges for the data collection may occur:

- (a) **There are limits on previous or ongoing monitoring data available** on birds, and funding challenges for carrying out special field studies. One has to bear in mind that the EBCC Pan-European FBI measures biodiversity at a continent or subcontinent level, whereas the need to measure the impact of RDP is a national/regional challenge in each MS. Therefore it is the obligation of the MS to care for the appropriate funding of this data collection. There might also be a need for cross-regional training, not in fieldwork as such but in setting down appropriate and most cost-efficient national schemes, and processing the results in a correct and relevant way. Lack of monitoring data and finances to carry out special studies may lead to inadequate or misleading evaluation results.
- (b) **The suitability of previous or ongoing monitoring data for evaluation of bird impact indicator** – in the case of using some previous or ongoing monitoring data for the evaluation of bird impact indicator the background and purpose of the data collected needs to be assessed to decide the suitability of the data. E.g. one problem could be that often areas with good environmental state are preferred – so, areas with not so good environmental state are not monitored. If such data is used, a wrong picture about the overall situation in the countryside might be reflected.
- (c) **Formation of monitoring sample** – the main weakness of most monitoring programs is not the actual counting method but the sampling strategy. The formation of monitoring sample is a challenging task. There are no identical farms – the landscape is different – the share of semi-natural habitats is different – farm sizes are different – etc. All these differences which

could affect the bird impact indicator results should be taken into account while forming the monitoring sample. Samples have to be representative in terms of habitat and geographical location. Therefore the best way to gather representative sample is to use large number of random plots. (<http://www.ebcc.info/wpimages/other/I4-SVENSSONH.pdf>)

- (d) **Comparison group** – There should also be a comparison group in the sample to find out what is the situation without the support, thus areas without the support should be involved. In some MS, e.g. Austria, Finland and Sweden the formation of comparison group for agri-environmental measures is practically impossible as they cover 90% and even more of agricultural land. In such cases monitoring results of different measures (e.g. broad and shallow measures versus specific species or habitat targeted measures) together with analysis of differences of measure requirements could form the basis for comparison.
- (e) **Data collection methodology** – There are several methods for bird data collection which have different quality in different circumstances.
It is important to use standardized methods as much as possible, so that the same method is always used at the same site every year. Results that come from a mixture of methods are equally as useful as results from a single method as long as a standard is maintained through time for each individual count.

5.1.5 Interpretation

Guidance Note J – Impact Indicator Fiches of the CMEF indicates that the evaluation could be conducted in two ways:

1. *“Estimated by programme evaluator at level of direct and indirect beneficiaries on the basis of output and result data, survey data and benchmark data and coefficients from similar projects and past evaluations (for calculation of double counting, deadweight, displacement). Cross-checking against counterfactual situation and contextual trends in programme area, particular as regards relevant driving forces, pressures and responses.*
2. *Estimation of contribution to general trend at programme area level (baseline trend), where feasible/statistically significant compared to other factors.”*

The following limitations should be considered when using birds impact indicators:

- ▶ birds are less specialized in micro-habitat use than many other taxa;
- ▶ their distribution at one scale may not match the patterns of other taxa;
- ▶ population trends may not always correlate with those of other taxa;
- ▶ environmental degradation can result in 'perverse' positive population trends in some situations;
- ▶ populations may respond to integrated sets of factors, rather than single ones, so their trends need to be interpreted with care. Agriculture is not the only factor impacting farmland birds, e.g. high density of predators may reduce productivity of birds and cause populations decline. In addition, climate change-related changes which may have considerable effects on farm and forest bird population, as well as the levels of disturbance and/or development should be considered. Furthermore, birds have the possibility to change their breeding site within the breeding season due to circumstances like drought etc.

5.1.6.1 Common baseline indicator no 17: Population of farmland birds

The FBI is designed to provide information about the general state of farmland biodiversity – and thus the environmental sustainability of agriculture – across the region or country to which each Rural Development Plan applies. Assuming a close link between the selected bird species and the farmland habitat, a negative trend signals that the farmed environment is becoming less favourable to birds and to biodiversity in general. For these reasons the FBI can be considered a good indicator of the impact of Rural Development Policy at European, national or regional level.

An assessment of the trends among breeding populations of characteristic birds can help to determine the quality of agricultural habitats and how this quality is changing through time. The negative trends of breeding populations indicate an unfavourable and worsening status of the bird species, which is very likely to be a useful proxy for biodiversity trends in general.

According to EBCC (<http://www.ebcc.info/index.php?ID=362>) the population trends could be classified and the changes in FBI values can be interpreted as follows:

- ▶ *Strong increase* – increase significantly more than 5% per year (5% would mean a doubling in abundance within 15 years).
- ▶ *Moderate increase* – significant increase, but not significantly more than 5% per year.
- ▶ *Stable* – no significant increase or decline, and it is certain that trends are less than 5% per year.
- ▶ *Uncertain* – no significant increase or decline, but not certain if trends are less than 5% per year.
- ▶ *Moderate decline* – significant decline, but not significantly more than 5% per year.
- ▶ *Steep decline* – decline significantly more than 5% per year (5% would mean a halving in abundance within 15 years).

The following challenges in interpreting the baseline indicator may occur:

- (a) **Different species may have different trends** – Farmland habitats bird specialised species give a good signal of biodiversity state over time as their sensitiveness is high. On the contrary, generalist species do often even benefit from habitat perturbations. Therefore, in a first step it is useful to check population trends of separate species to identify whether any of those included in the list may have had rapid changes in populations during the period under consideration. These species may influence the FBI considerably (e.g. species with opposite trends may compensate each other), as FBI show the average trends in abundance of a selected set of species. They are especially useful in showing change in the overall condition of ecosystems, which is difficult and expensive to measure directly. So, if the indicator is not doing well, the indicator should be analysed to see which species is/are causing the decline, and initiate research on these species (e.g. study the reproductive success, distributional changes, etc.)

Effects of trends of individual species on the overall index performance – an example from the Netherlands

In The Netherlands, 25 years ago with a common bird census (CBC) like method (more info about CBC: <http://www.bto.org/survey/complete/cbc.htm>) for all breeding birds was started. Based on the nation wide coverage of this program, all kind of selected CBC subcensuses were formed, like CBC meadow- and farmland birds, CBC rare birds, CBC colony breeders etc. At this moment there is also a wish for a more

specific program, called MAS (monitoring agricultural species). To enhance the knowledge of the breeding success, the work with alarm counts on Black-tailed Godwits is carried on. Results of all these efforts are reported every year. After 25 years of monitoring, it was concluded that within species groups different trends occur: most meadow birds are declining, but some are increasing, some migrants are decreasing due to the situation in the wintering areas, some are increasing (also) due to better breeding conditions in The Netherlands.

And why Tree Sparrows and White Wagtails are declining in the Netherlands, is still the question. But it should have something to do with agricultural landscapes. Problem is mostly: what is the key to success? Montagues harriers had the best breeding season since 30 years with 60 breeding pairs and a good breeding result. But: it was a bad year for the common vole, the bulk feeding species for this raptor. Due to a detailed mosaic management in both natural and agricultural areas, this species is doing very good, so maybe the key problem for it was found.

Data source: Comments of Cees Witkamp, Rene Alma, DLG Government Service for Land and Water Management

A good practice could be to exclude those species that show wide fluctuation over time to avoid that very few species influence greatly the Index. Those species are usually also quite rare (not widespread, small numbers) and subjected to stochastic fluctuation. For example, Italy excluded species that occur in less than 10% of samples.

The Index should contain species that depend on different farmed habitat. It can be useful to check population trends of groups of species with similar habitat requirement separately (e.g. meadow birds and hedge birds). These can also be analysed in relation to the provision of resources within a life-cycle context. Most species need three main requirements: nesting habitat, chick food and habitat to support overwinter survival. For resident species all three may need to be provided effectively to sustain and/or enhance populations. It could happen that different group have different trends that are not detectable in the Index as a whole.

- (b) **There are a number of environmental factors that influence the FBI contributing to the recent decline** – a key factor however is considered to be the area of suitable habitat such as the area of un-cropped and other fallow land. Disease has had an important influence for some species, such as greenfinch, where there has been a particularly sharp fall (-15%). In case such incidents are known in a country it would be wise not to include affected species into the set of species used for index.

Individually, many species may show annual changes in abundance that may reflect a variety of environmental factors, such as extreme weather conditions during the breeding season, poor conditions on the winter grounds, changes in predation pressure, and simple sampling error and statistical noise.

Use of multivariate impact assessment for filtering out environmental factors that influence the Farmland Bird Index – an example from Finland

Farmland bird monitoring is conducted with a mapping method in different parts of southern Finland. Between 59–141 study plots totalling ca. 6,100–12,700 hectares of farmland have been censused in 2000–2009 (one year with smaller coverage). All the territories representing almost 50 species have been stored in a GIS database. These data are then used for assessing impacts of agri-environmental measures and many other factors influencing bird numbers with spatial modelling methods in appropriate scales. Explanatory variables are drawn from different spatial databases, like the field register providing information of field use annually, and vectors from basic maps of the General Land Survey of Finland.

Data sources: Finnish Game and Research Institute, Juha Tiainen

- (c) **Periods of time for monitoring FBI.** FBI is **indexed** on the year 2000 but Member States may choose other years if this improves the quality of the analysis. In fact it is better to use longer periods if possible. It is not very difficult to change the base year in analysis (in

TRIM). While looking at different base years, the change compared to these base years is easy to understand. Another aspect is to 'synchronize' the index with agri-environmental programmes (often there are 'system changes').

5.1.6.2 Impact indicator – change in trend in biodiversity decline as measured by farmland bird species population

Depending on the extent of specific management of farmland under agri-environmental schemes responses of changes in bird populations may have different time lags (sometimes it may take several years before management effects will become obvious). It has to be considered that birds will take time to respond to measures provided by AES, especially in case of broad and shallow agri-environmental schemes, which may not meet bird feeding and nesting needs and therefore impact might be not measurable even after some years!, as suggested

It should be also considered that other factors e.g. variations of crop rotation may influence monitoring results. In latter case edge-specialist species should be analysed together with field species as they are less dependent on changes in crop rotation. Improved use of existing measures through scheme targeting, advice, and correct location of scheme options, can deliver a recovery for farmland birds.

The interpretation of the impact of agri-environmental measures on biodiversity demands that several factors which could have affected the results will be taken into account in data analysis. It helps to find out the net effect of the agri-environmental measure.

Often detailed and specific field studies are necessary to identify precisely the contribution of AE measures. Detailed field studies with a careful sampling design and statistical analysis cannot only contribute to the evaluation of AE measures, but can be used furthermore to set realistic and concrete quantitative conservation targets, to develop 'optimal' schemes and to estimate costs while securing cost-effectiveness.

One main weakness of field studies is that in most cases they look at specific situations; the selection of study sites can be biased, and in such cases they are not representative for the AE measures implementation at national (or other relevant) level.

However as the following box clearly shows, Austria has been implementing a wide range of specific studies over a long period of time, which has not only provided a series of important specific findings in relation to AES but also enabled the MA to design and set up more efficient monitoring systems for the present 2007-2013 programming period; specific studies have therefore been streamlined and closely interrelated.

Assessment of Impacts – Interpretation of agri-environmental measure effects using multiple sources:

In order to evaluate AE measures' effects on biodiversity within the RDP 2000-2006 period, Austria used data from the Common Bird Monitoring scheme, which is based on yearly point counts in Austria.

A total of 38 farmland bird species were analysed. By GIS tools, for each of app. 1100 farmland count points, a 'buffer' plot was created; these plot varied in size depending on the species studied (range: 50 to 500 meters radius). For each plot, a large number of variables was created (e. g. area percentages), which included not only the whole set of AE measures, but even land use (e. g. type of cereals, field size), landscape (e. g. forest cover), and other 'control variables' like soil productivity or altitude. The presentability of these monitoring points was tested by comparing them with plots randomly spread over the whole of Austria, and no significant differences were detected.

Each of the 38 species was analysed separately by multivariate statistics. Logistic regression was used to investigate if and to which degree the presence (or absence) of a species is depending on AE measures and/or agricultural land use, while controlling for other influences (e. g., soil productivity, altitude).

It emerged that most of the species are influenced positively by at least one AE measure, but the relative contribution varies greatly, and in general is rather small compared with other factors. One important finding was that even some of the 'light green measures' may have beneficial effects on farmland birds, but the effects are very small compared with the spatial coverage of those measures. On the other hand, well targeted measures (e. g. those in the 'conservation scheme'), may have significant effects even if implemented on a small proportion of agricultural land. Therefore, this study could demonstrate clearly that the impact of AE measures on the population level is depending both on the 'strength' (prescriptions) as well as on the scale of implementation (uptake).

A similar multivariate study based on bird monitoring data, farming and measures under RD will be done in 2010, aiming again to identify the net effects of the Austrian AE scheme on birds (both spatial and temporal).

Data source: Frühauf & Teufelbauer 2006.

The use of detailed field studies for closing the cause-effect gap

Detailed field studies with a careful sampling design are necessary to identify precisely the contribution of AE measures. It is essential that a wide array of 'control variables' is included in the analysis and that multivariate statistics are applied if 'net effects' of AE measures are to be evaluated. For example, when areas with higher biodiversity are detected, it is crucial to be able to distinguish between areas with low productivity and the effects of the AE measures itself (which often are implemented only on poor soils). The results of such studies can be used to improve AE schemes.

In the RDP 2000-2006 period, in Austria a very detailed field study was done for the evaluation of AE measures effects on biodiversity in arable land.

The investigations focused on four target species which differ in foraging and breeding ecology and included migratory as well as not migratory species (Grey Partridge, Quail, Skylark, Brown Hare). By intensive field work, densities of all four species were quantified within 92 study plots of 28 ha each, as well as reproductive success for Partridge and Hare; furthermore, the number of farmland bird species registered in each plot was used for analysis.

For each study plot, a large number of variables was created (e. g. area percentages), which included all AE measures, land use (e. g. type of cereals, field size, diversity of land use), landscape (e. g. tree lines), data of individual farmers (including, e. g. machinery, details of pesticide use and set aside management, size of holdings), vegetation of set asides (plant species richness, vegetation structure), and other 'control variables' like soil productivity, distance to roads or pressure from predators. Multivariate statistics were used to identify the effects of AE measures and to control for all other influences.

The study produced many important results. For example, it emerged that set asides implemented under the AE 'conservation scheme' with specific prescriptions have by far larger influence on the settlement of partridge territories than the much more numerous 'normal' set asides. The study produced many insights regarding the effects of organic farming. Cereal fields under the organic scheme were more attractive for all species investigated than conventional fields under 'light green schemes', as the former provide much more food (herbs, seeds, invertebrates). However, as organic farmers are not obliged to set aside, they do not contribute to reproduction; moreover, alfalfa, which is cultivated by organic farmers to secure the fertility of soils, is very attractive for all species, but is a 'sink' because of improper management (too early and too frequent cutting destroys nests and kills young).

The timing of the management of set asides resulted to be crucial for the reproductive success in several of the species investigated. No negative effects of predators on reproduction success could be detected. In general, farmland bird species richness is positively correlated with set asides and small fields, and negatively with intensive farming (e. g. use of pesticides, intensive cultivations as sugar beets).

Several concrete recommendations out of this study were integrated in the current AE organic scheme (2007-13).

Data source: Kelemen & Frühauf 2005.

The experience from these examples from the previous programming period has lead to a comprehensive monitoring scheme and helped to set realistic conservation targets

Detailed field studies with a careful sampling design and statistical analysis cannot only contribute to the evaluation of AE measures, but can be used furthermore to set realistic and concrete quantitative conservation targets, to develop 'optimal' schemes and to estimate costs while securing cost-effectiveness.

In 2008, in a mountain (800-1400 m a.s.l.) grassland area in western Austria a very detailed field study was done on the effects of AE measures on the Whinchat, a threatened and decreasing meadow breeder.

Field data collection included the localization of 82 breeding territories and the assessment of their breeding success

(fledged young) as well as the mapping of all potential hunting and singing perches. By GIS work, several relevant AE measures, land use (e. g., meadows cut one to three times a year), perches, productivity, altitude etc. were related to 'ideal' (circular) territories and equally sized random plots. By multiple analysis (logistic regression), the presence of Whinchat territories, and of successful territories could be explained by AE measures and the other variables included in the analysis (e. g., the availability of the preferred small-sized perches).

A major result was that 'light green measures' may contribute to the establishment of territories and to spatial distribution of feeding Whinchats, but only the targeted measures under the 'conservation scheme' (which include, e. g., late mowing) had a significant effect on breeding success, although they cover only a very small percentage of the study area.

In a second step, using the results of the logistic models, it was explored (by modelling) which amount and combinations of specifically designed conservation measures would be needed to increase substantially the number of successful territories under realistic assumptions. For a realistic increase (over 90%) an optimal 'strategy' (including costs) was modelled. The careful design of this study allowed to show that the area percentage under specific management (late cutting of meadows) necessary for achieving a population increase in the Whinchat is much smaller than it was stated in other studies.

Data source: Peer & Frühauf 2009.

5.1.6 Judgement issues

The approaches for assessing and evaluating impacts in the field of biodiversity decline are to be assembled for the final judgement on the RD programmes' impacts. This means that all sources of information will have to be taken into account, with the previous chapters pointing out several approaches, which may be adopted by evaluators.

Moreover, when an additional objective or national priority defined in the programme is not covered by an impact indicator or the common impact indicator is not detailed or specific enough to reflect the wider benefits of a measure, or where a common impact indicator does not exist for a measure it is recommended that MS should provide additional appropriate impact indicator. This is particularly important where measures are highly site-specific, for example in agri-environment.

The definition of additional indicators will give Member States flexibility in creating a monitoring and evaluation system adapted to their needs.

- ▶ One complementary indicator which could be used in capturing programme impacts, is to look at the **share of farmland birds with declining populations**. The main indicator (FBI) and the current sub-indicator are based on different data sets; the sub-indicator does highlight and confirm results of the main indicator. Whereas the main indicator shows the trend of the populations of farmland bird species collectively, the sub-indicator shows proportion of species with declining populations. Both indicators refer to the same list of farmland bird species. Data for the sub-indicator come from national estimates of the overall trends in bird population sizes from 1970 until 1990 (BirdLife International/EBCC, 2000). Population trends are provided as classes:

- Decreasing: population decrease of at least 20%
- Stable: population with overall change less than 20%
- Fluctuating: population fluctuating with changes of at least 20% but no clear long term trend
- Increase: population increase of at least 20%

Similar complementary approaches should be explored for forest related measures.

The use of additional impact indicators – an example from England

There is a supplemented impact indicator in England “Reversing biodiversity decline: the change in trend in biodiversity decline as measured by farmland and woodland bird species population”. The following farmland bird species will be monitored to measure against this indicator: tree sparrow, corn bunting, grey partridge, turtledove, reed bunting, starling, skylark, linnet, lapwing, yellowhammer, yellow wagtail, kestrel, goldfinch, greenfinch, wood pigeon, jackdaw, stock dove, whitethroat, barn owl, and rook.

The woodland bird population index will be used in addition to the farmland bird index. 33 woodland bird species are included in the index, 12 of which are woodland generalists and 21 are ‘woodland specialists’. The target for this indicator is to reverse the long-term decline by 2020. The figure will be updated annually based on the Common Bird Census and the Breeding Bird Survey. The impact of the program on this indicator will be assessed qualitatively using sample-based monitoring and research.

Data sources: <http://www.defra.gov.uk/corporate/consult/gaec/info-note-farmlandbirds.pdf>
<http://www.defra.gov.uk/rural/rdpe/progdoc.htm>

- ▶ It is also useful to further **explore population trends of individual bird species** that are reported by the Pan-European Common Bird Monitoring project (time period 1980 to 2006). Individual species trends help to understand mechanisms that drive trend in farmland birds. It is easier to link certain support schemes to individual species or guilds rather than to a large group of species with different or even contradicting diet and habitat selection preferences. In case bird data are not representative for all species of interest more numerous species with specific needs may be taken as an example for analysis.

5.1.6.1 Conclusions

- ▶ **Data availability, reliability, representativeness and regional suitability on birds for baseline (FBI) and impact indicators should be found out as early as possible**
- ▶ **MS should use an alternative composition of bird species for FBI where this is appropriate to national/regional situation.** Meanwhile, it is crucial to state which indicator is used for which purposes because indices with different species composition indicate different trends.
- ▶ **Additional analyses with FBI on a smaller scale.** In a first step it is useful to check population trends of separate species to identify whether any of those included in the list may have had rapid changes in populations during the period under consideration. There is also a possibility to subdivide the FBI into several classes to allow deeper insights
- ▶ **Other indicators, than birds can help to prove positive impact of the AE scheme.**
- ▶ **The formation of the monitoring sample is a challenging task.** Samples have to be representative in terms of habitat and geographical location. Therefore the best way to gather representative sample is to use large number of random plots.
- ▶ **Counterfactual situation, net effect and the need to apply more complex statistical methods (for reducing biases).** In most cases, the evaluation of individual measures or schemes requires ad hoc and highly replicated field studies, including pair-wise comparisons with control sites. Evaluations must take place over an appropriate length of time (dependent on species/community being evaluated).

- ▶ It is very important, however, to **‘control’ other influences**, which is possible by careful design and would need multivariate analysis in most cases. It means also taking into account environmental variability that is not directly linked to the measure and partial it out, making it possible to separate effects actually caused by the measure on bird populations from other spurious effects.
- ▶ **Establish the chain of evidence between results and impacts.** It is crucial that there is a real link between biodiversity and ‘successful’ land management. The MS should be able to show which AE measures or types of land use (both cropped and uncropped) are really contributing to the improvement of biodiversity and to which degree. Thus, **additional studies on management practices may be necessary**. Evidence based management options also help to improve the measures.
- ▶ **The quantities measured have to be put into a general context.** The meaning of the evaluation results (e.g. positive and negative trends of the indicators) has to be put into a general context to be understood.

One crucial factor in assuring the data availability which is necessary for evaluation AE measure is **financial constraints**. In that case better cooperation and harmonization of monitoring activities of agricultural and environmental policies in the country should be encouraged and combination of national and EU financial tools (e.g. technical assistance measure under RDP) supported. Lack of monitoring data and finances to carry out special studies may lead to inadequate or misleading evaluation results.

5.2 Impact Indicator 5: Maintenance of HNV farming and forestry

This chapter on HNV farming and forestry complements the Guidance Document on the Application of the High Nature Value Impact Indicator, now incorporated into the CMEF Handbook⁸¹. In particular, aspects concerning the establishment of the baseline situation and monitoring activities are covered in this document and should be used as additional guidance. As the HNV common impact indicator is new in the programme period 2007-2013 compared to previous periods, this chapter provides important guidance on acceptable methods for defining the baseline situation – a challenge which Member States must address before contemplating methods for measuring the impact of RD programmes on HNV farming and forestry.

5.2.1 The CMEF guidelines

5.2.1.1 The common impact indicator: changes in High Nature value Farmland and Forestry

According to guidance note F of the CMEF, the impact indicator in this case is:

“Changes in high nature value *farmland and forestry*”.

Guidance note J of the CMEF interprets the impact indicator as follows:

“*Changes in high nature value areas*”

The relevant policy priority as set out in the Community's Strategic Guidelines for rural development is to use measures to preserve HNV farming and forestry systems.

How should Member States interpret these different HNV terms?

HNV farmland refers to farmland characterised by the presence of particular land cover types and patterns (especially semi-natural vegetation and low-intensity crop mosaics) which indicate that this farmland is valuable for nature conservation. The presence of populations of particular wildlife species may also provide this indication. HNV farmland may exist at different scales, from the individual parcel to an entire landscape.

HNV farming system refers to both the land cover (farmland) and the way it is managed for production by a particular farming system and practices. The term implies that the system as a whole (e.g. at farm or even landscape level) is of high nature value, whereas HNV farmland may be limited to only one parcel in an otherwise intensive farming system.

The same interpretation can be used for the terms HNV forest and HNV forestry system.

The terms HNV farming and HNV forestry are used in this document to refer to the overall concepts without distinguishing land from management system.

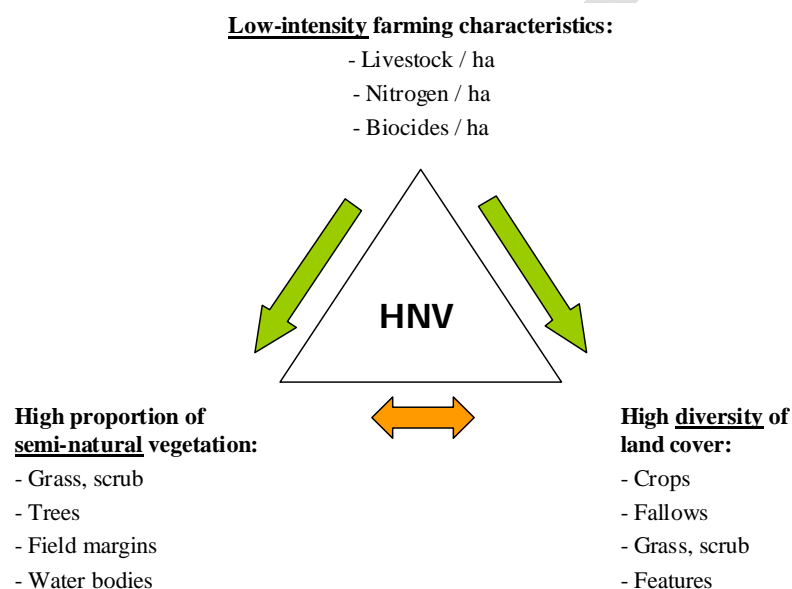
⁸¹ http://ec.europa.eu/agriculture/rurdev/eval/hnv/guidance_en.pdf

The introduction of the term *areas* in guidance note J has led to some confusion. The HNV Guidance Document (EENRD, EC, 2008) emphasises that the idea of the indicator is not to *designate* particular *areas or zones* as HNV. However, in practice, particular types of HNV farming and forestry may be concentrated in certain zones, and it may be useful to identify approximate zones as a practical basis for establishing appropriate indicators for monitoring tendencies within distinct zones.

The idea of the HNV concept is to contribute to nature conservation by supporting and maintaining the broad types of farming and forestry that favour biodiversity, because of their *characteristics*. The HNV Guidance Document explains the broad farming and forestry characteristics that are known to be critical for supporting nature value, and which then provide the basis for identifying HNV farming and forestry on the ground. Figure 7 summarises these characteristics for HNV farming.

As the diagram illustrates, high nature value results when certain patterns of land cover (those with a high proportion of semi-natural vegetation and a diversity of types) are managed for production in a particular way (under low intensity systems and with particular farming practices).

Figure 7 The Three Key Characteristics of HNV Farming Systems



Three types of HNV farming have been defined:

- ▶ Type 1 – Farmland with a high proportion of semi-natural vegetation.
- ▶ Type 2 – Farmland with a mosaic of low-intensity agriculture and natural and structural elements, such as field margins, hedgerows, stone walls, patches of woodland or scrub, small rivers etc.
- ▶ Type 3 – Farmland supporting rare species or a high proportion of European or World populations.

Farmland of the first type is generally very species-rich, by definition requires extensive agriculture for its maintenance, and has well-recognised conservation value. The second type is defined because small scale variation of land use and vegetation, combined with low agricultural inputs, is generally associated with relatively high species richness. A smaller proportion of the farmed habitats within this type will be strictly semi-natural, but the management should be sufficiently extensive to allow for floristic variation.

The third type is defined because locally more intensive farming systems which do not have the characteristics of Types 1 and 2 may sometimes support significant populations of species of conservation concern.

It should be stressed that the term HNV is applied to farmland or forest where the productive area itself (grazed vegetation, crops, tree stands) supports significant biodiversity. Normally this occurs in agriculture when a significant part of the land cover is under low-intensity use and is thus in, or close to, a semi-natural state. More intensively farmed land is only HNV in limited circumstances where it supports certain species of conservation concern, normally bird species (Type 3).

Where semi-natural landscape features, such as hedges, ponds and small uncultivated patches, survive on intensively managed farmland that otherwise is of limited nature value, these features are important for conserving vestiges of biodiversity. They may be considered as HNV landscape features, but by themselves these features do not qualify intensively managed farmland as HNV farmland.

In the same way, it may be possible for non-HNV forests to contain HNV features, if these support significant populations of species of conservation concern.

The overall challenge for Member States in order to implement this CMEF indicator is to:

- ▶ Devise a set of indicators that will provide meaningful information on changes in the *extent* and in the *condition* of HNV farmland and forests, and on trends in HNV systems and practices, during the seven years of the rural development programmes.
- ▶ Devise a method for assessing to what extent (and how) these changes and trends have been influenced by RD programmes and measures.

This chapter focuses mainly on the first of these two steps, as many Member States have not yet established a satisfactory method or set of indicators for this. There are many challenges still to overcome. This is a fundamentally different situation from that of, for example, the Economic Growth indicator – whereas there are established and accepted methods and data sources for measuring Economic Growth, this is not yet the case for HNV farming and forestry.

It is impractical to contemplate methods for measuring the impact of RD programmes on HNV farming and forestry without first establishing an acceptable method for measuring the baseline situation. This requires an analysis of HNV farming and forestry that goes beyond mapping exercises. Thus, the CMEF guidelines recommend that the ex-ante evaluation should include an assessment of:

the handicaps facing farms in areas at risk of abandonment and marginalisation;

overall description of biodiversity with focus on that linked to agriculture and forestry, including high nature value farming and forestry systems

These can be seen as part of a qualitative approach to establishing the HNV baseline situation (see below).

5.2.1.2 The Common Evaluation Questions touching HNV Farming and Forestry

The Common Evaluation Questions in Guidance Note B do not refer specifically to HNV farming and forestry, although for Axis II there are several questions asking to what extent measures have

contributed to the maintenance of “sustainable farming systems”, “the countryside”, “landscape and its features”, and “biodiversity”.

For all measures, it would be appropriate to consider the question: To what extent has the measure contributed to maintaining or improving the economic viability and/or ecological condition of HNV farming and/or forestry systems? (Note that measuring economic viability is not a specific CMEF indicator for HNV, rather it makes a link to the ex-ante evaluation of handicaps faced by farms in areas at risk of abandonment).

The relevant Horizontal Evaluation Question is:

To what extent has the programme contributed to promoting sustainable development in rural areas? In particular, to what extent has the programme contributed to the three priority areas for protecting and enhancing natural resources and landscapes in rural areas:

- ▶ biodiversity and **the preservation and development of high nature value farming and forestry systems** and traditional agricultural landscapes?
- ▶ water?
- ▶ climate change?

Methods for evaluating programme contributions in the areas of biodiversity, water and climate change are addressed in other chapters of this document.

5.2.2 Key challenges with regard to measurement and interpretation

Logically, the first step is for each Member State to assess the baseline situation against which the changes in HNV farming and forestry can be measured.

The baseline indicator is defined in quantitative terms in the CMEF guidelines as “UAA of HNV farmland”. There are two points of clarification to make with respect to the CMEF wording:

1. Although not specified, the baseline indicator logically should cover HNV forests as well as HNV farmland.
2. The term “UAA” should not be taken as covering only land within the farm holding, as is the case with the Farm Structures Survey (FSS). Farmland off the holding, especially grazing land in common usage, should be included in estimates of the baseline extent of HNV farmland. Such off-farm grazing constitutes a large proportion of HNV farmland in some regions.

In order to explore the most appropriate tools for calculating the extent of HNV farmland and forests, each Member State needs to clarify what types of HNV farming exist in its territory, by interpreting the EU definitions of HNV farming (see HNV Guidance Document⁸²) in the national context with input from multi-disciplinary expert knowledge. This is an essential part of establishing a qualitative baseline.

⁸² http://ec.europa.eu/agriculture/rurdev/eval/hnv/guidance_en.pdf

Establishing a qualitative baseline

As explained below, there are difficulties at present with a quantified approach to establishing the baseline (i.e. number of HNV hectares). It is important therefore to complement the quantified estimate with a *qualitative assessment of the baseline situation of HNV farming and forestry*. This should be based on available information and expert knowledge and aim to provide a practical and descriptive picture of:

- ▶ the types of HNV farming present in the programme area, and their approximate distribution
- ▶ the main characteristics and practices of the farming systems and how these are related to specified biodiversity values
- ▶ the socio-economic situation of HNV farming types and main challenges faced, and current perceived tendencies (in order to provide an insight into trends to be expected in the counterfactual situation, such as abandonment and/or intensification).

The baseline evaluation should give sufficient attention to assessing the situation of HNV farmland along these lines, for example through multi-disciplinary studies and working groups. It should also propose quantified objectives and targets which can then form the basis for evaluation of the programme and its components, as described in the CMEF guidelines.

Many Member States have not provided this qualitative assessment of their HNV farming and forestry baselines in the 2007-13 RDPs. This needs to be addressed as soon as possible, otherwise it will be extremely difficult, if not impossible, to evaluate the impact of RD programmes on HNV farming and forestry. In the case of the HNV indicator, the mid-term and ex-post evaluations will depend on a considerable qualitative element, and therefore a qualitative baseline is needed against which to assess changes during the programme period.

In practical terms, the data gathering and interpretation for establishing a qualitative baseline is best undertaken alongside the exploration of data for the quantitative baseline, as described below.

Approaches to quantifying the HNV baseline

Member States have been addressing the challenge of estimating the ex-ante extent of HNV farming and forestry, as part of the baseline indicator.

HNV farming has certain characteristics that are known to favour biodiversity. As shown in the diagram above, these characteristics can be divided into:

- ▶ **Land cover** characteristics, especially farmland with a high proportion of semi-natural vegetation and in some cases a diversity of land cover types.
- ▶ **Farming practices**, especially a low use of inputs (including livestock density) and specific practices such as shepherding, late hay-cutting, orchard grazing and arable fallowing.

The term HNV farming system conveys the principle that it is the interaction of *farmland characteristics* and associated *farming practices* that produces a situation of high nature value. This is important in the context of the evaluation of rural development programmes, since measures impact on farming practices, and via these, also on the land itself.

In the same way as HNV farming, HNV forestry is defined by a combination of two types of characteristic:

- ▶ **Land cover characteristics**, namely the composition and condition of the forest vegetation, including the under-storey. Important characteristics include species composition, age classes, presence of mature trees and deadwood.
- ▶ **Forestry practices**, especially forest management practices that favour biodiversity.

An important difference between HNV forestry and HNV farming is that whereas HNV farming consists of semi-natural habitats that depend by definition on human intervention for their maintenance, some types of forest are of high-nature value without needing human intervention. These natural, rather than semi-natural, forest types can be defined purely by their physical characteristics.

Semi-natural forests require sustainable forest management methods to maintain their semi-natural state. This may include in some cases grazing and browsing by livestock. In these cases, there is an overlap between HNV farming and HNV forestry that needs to be taken into account when defining systems and indicators.

Under an ideal approach to identifying the location and extent of HNV farming and forestry systems, data sources would be used to identify land where certain land cover types and certain management practices coincide. To be accurate, this would be done at the level of the farm or forestry holding, or even at the parcel level. Thus, for a given region, by crossing data on land cover types and on farming/forestry practices, it would be possible to identify the land units that exhibit a chosen set of HNV characteristics, although it is not the intention to classify farms as HNV or non-HNV.

In practice, there are severe data limitations at present which make this approach very difficult at the level of entire countries or regions. The data required are not available to distinguish the full range of HNV characteristics at the level of a land parcel or farm/forest holding, or to map their distribution with accuracy across an entire region.

Faced with this situation, Member States are using approaches to identifying the baseline extent of HNV farming (and in some cases forestry), using various types of data and methods. These are discussed below, according to main approaches used and potentially available:

- ▶ Land cover data
- ▶ Farming systems data
- ▶ HNV “scores” based on mixed criteria
- ▶ Species distribution data
- ▶ Sample surveys

5.2.3.1 Land cover approach

Key types of land cover that make up HNV farmland include:

- ▶ Semi-natural pastures and hay-meadows.
- ▶ Traditional orchards with semi-natural under-storey.

- ▶ Mosaics of low-intensity crop types with a high density of semi-natural landscape elements, such as field margins and patches.
- ▶ Fallow land in low-intensity arable systems, when spontaneous vegetation is allowed to develop.

In the case of HNV forests, the broad land cover types are natural and semi-natural forests.

Several Member States have used CORINE in an attempt to estimate the extent of relevant land cover types, following the methodology of JRC/EEA. CORINE land cover classes are selected that have a high probability of coinciding with HNV farming. For example, Bulgaria, Greece, Hungary and Romania have followed this approach.

CORINE data represent land cover and not land use. The data therefore do not contain information on the intensity of management (e.g. input use, grazing pressure). Furthermore, the level of detail of the map (25 ha of minimum mappable area) does not allow the identification of small, scattered surfaces which may contribute to high nature value. JRC/EEA have used CORINE for a harmonised approach at the EU scale, supplementing the data as far as possible with biodiversity data and national inventories. However, it is stressed that this approach provides only a proxy distribution of HNV farmland in Europe and is not intended nor suitable for evaluating the impact of rural development measures at national or regional level.

Semi-natural pastures and meadows

Semi-natural pastures and meadows consist of native species that have not been artificially sown or fertilised. Pastures can include grassland, scrub, woodland, or a combination of these types. Scrubby and woody pastures can be of particular importance for biodiversity conservation and represent a widespread type of HNV farmland in some regions.

CORINE is relevant mainly for identifying pastures that are likely to be in a semi-natural state. Two land cover categories are most relevant – Natural grassland (3.2.1) and Pastures (2.3.1). There are difficulties with the second category in particular: although likely to include semi-natural hay meadows and some semi-natural grazing, it also includes more intensively used grassland that may be resown and heavily fertilised.

For other land cover types that may be used for low-intensity grazing (e.g. Moors and heathland 3.2.2), CORINE is not able to distinguish areas that are used for domestic livestock grazing, from areas that have no real farming use.

CORINE therefore gives a very approximate “picture” of the extent of vegetation that may be under HNV livestock farming, but this is likely to include areas of intensive grassland, and areas of non-grazed scrubland. The result is too crude to be taken as an acceptable estimate of the extent of HNV livestock farming.

One approach is to enrich the background picture provided by CORINE in various ways using national data.

Example of enriching CORINE land cover approach – Bulgaria

In Bulgaria several land cover data were added, including:

- Types of Natura 2000 habitat threatened by abandonment of extensive agricultural practices (mainly grazing),

- from national inventory of Natura 2000 habitats.
- Semi-natural grasslands, from national inventory of semi-natural grasslands.

These data have been incorporated in the LPIS so that blocks of parcels can be identified as HNV using this system.

Some countries have well-developed national inventories of semi-natural grasslands and meadows, which may capture a large proportion of HNV livestock farming with considerably more accuracy than CORINE.

For example, in Sweden the national Inventory of Valuable Pastures and Meadows provides a complete picture of the distribution of these land cover types. Similarly for the approach being developed in England, land cover data from the Countryside Survey provide the distribution and extent of the main types of semi-natural vegetation that may be under farming use.

The Land Parcel Identification System (LPIS) is potentially useful for identifying types of pasture that are more likely to be in a semi-natural state, for example, categories such as rough grazing and pasture consisting partly of scrub and/or trees. However, caution is required as the broad category of permanent pasture as defined under the CAP includes more intensively used grasslands, and these may be lumped together in some countries.

Also, pastures at the most natural end of the spectrum, with a dense cover of scrub and trees, are not counted as farmland on LPIS in some countries, even when such land cover is grazed and is highly likely to be HNV.

Orchards

European data bases such as CORINE and FSS do not distinguish between intensive and non-intensive permanent crops, so are of almost no use in this case. The only distinction is between irrigated and non-irrigated crops, for example in the case of olives in FSS.

At a very general level, non-irrigated permanent crops in Mediterranean regions are more likely to be close to a semi-natural state than irrigated crops. Thus at a first level of identification of HNV farming in Greece, non-irrigated olives are taken as potentially HNV. Ideally, this criterion should be complemented with others reflecting variables such as mosaic patterns (e.g. diversity of land cover types, modal size of parcels), and presence of semi-natural elements in the landscape (see below).

As with a category such as permanent pasture, a simple decline or increase in the regional extent of non-irrigated olive plantations would not be a reliable indication of tendencies in HNV olive farming. Tendencies in management practices, such as input use or allowing the growth of a spontaneous understorey at least for part of the year, are likely to be more significant for biodiversity than is a change in the overall extent of non-irrigated olives. Relevant management practices probably can be monitored only through sample surveys (see below).

Some countries have specialised inventories of traditional orchards, for example France, in which case these can be used to monitor trends in this specific type of HNV land cover.

Mosaics of low-intensity crop types with a high density of semi-natural landscape elements, such as field margins and patches

Some countries have experimented with methods for identifying areas of farmland with a concentration of small fields and/or high diversity of landscape elements, for example Sweden.

Mosaic patterns and the presence of small fields and linear elements generally would not be taken as an indication *by themselves* of HNV farming. It is important that a high proportion of the land cover should be under low-intensity uses, but this is difficult to ascertain from existing land cover data.

In some regions, it may be appropriate to assume that certain crops are more likely to be cultivated at a low intensity, such as rye, oats, and non-irrigated Mediterranean tree crops. LPIS may be useful for identifying areas where a concentration of such crops exist in mosaics.

For example, a study by the Spanish Government used data extracted from LPIS on heterogeneity of parcels at the holding level to include in a multi-variable methodology for identifying HNV areas.

The presence of semi-natural landscape elements is an important HNV criteria in the case of crop mosaics. Larger elements, such as patches of scrub or woodland, may also be detected through LPIS, though smaller elements (hedges and headlands) may be hard to identify.

Arable fallow

HNV farming of arable crops is only found under traditional, low-intensity systems. In southern Europe, a key characteristic of HNV arable cropping is a high proportion of fallow. These land cover types cannot be identified from CORINE.

However, arable fallow is recorded on FSS and the extent of fallow land at municipal and regional level also can be extracted annually from IACS data. The extent of arable fallow land, for example at the municipality level, is an important background indicator of tendencies in HNV farmland in parts of southern Europe. It is one of very few indicators that is captured relatively easily from existing data sources.

Forests

Forest inventories often provide more information than other land cover data, and may enable the estimation of a baseline extent of HNV forests.

The Sustainable Forest Management monitoring system, introduced by the Ministerial Conference on the Protection of Forests in Europe (MCPFE), includes biodiversity indicators that are useful for the identification of HNV forests. In the following example from Italy, the MCPFE indicators are compared with attributes from the national forest inventory (INFC) that can be used for this purpose.

Example of HNV forest baseline indicator – Italy

The main source of updated data on Italian forest resources is the recently completed second Italian NFI (2003-2007, reference year 2005), called “National Inventory of Forests and Forest Carbon Sinks” (INFC). Data collection was performed on sampling points selected on the basis of a three phase sampling strategy. The advantages of using INFC data are the following:

- (a) Statistical sampling and homogeneous analysis throughout the Italian territory;
- (b) Information consistent with international standards (FAO and MCPFE);
- (c) Detailed analysis of certain aspects of forest biodiversity, for example thanks to a detailed forest classification;
- (d) Possibility of revising the inventory.

Potential disadvantages include:

- (a) The data referring to single points cannot be extrapolated to areas: therefore, it is possible to estimate precisely the extension of national and regional HNV forests, but not their location (mapping of HNV forest areas can be done by the competent regional bodies);
- (b) Uncertainty concerning INFC repetition timing.

MCPFE indicators and INFC attributes to estimate HNV forest areas

MCPFE indicator	INFC attribute	Possible use for the estimation of HNV forest areas
Naturalness (4.3)	Forest category	Makes it possible to distinguish natural and semi-natural forests from plantations.
Introduced tree species (4.4)	Forest sub-category	Makes it possible to distinguish forests made up prevalently of autochthonous species.
Regeneration (4.2)	Origin of the stocking	Makes it possible to distinguish natural or semi-natural forests from reforestation and forestation
Protected forests (4.9)	Protected areas	Meets the criterion of presence of species that are particularly valuable for the preservation of biodiversity
Dead-wood (4.5)	Total dead-wood	The volume of dead-wood is one of the parameters advised by the EC. It is possible to draw some information about this indicator from the INFC recording relating to stand development stage.
Specific composition (4.1)	Number of tree species	Tree diversity, though not a parameter advised by the EC, is deemed of considerable importance. It is possible to draw some information about this indicator from the INFC recording relating to the forest sub-category.

In the light of these elements, at the moment it is deemed necessary to supply an estimate of HNV forest areas based on the processing of more general information – of a qualitative nature – collated by the INFC. A method test, performed on INFC data, has led to classify as HNV forests approximately one fourth of Italian forests. In future years, if deemed necessary and on the basis of possible new data (e.g. INFC revision, ad hoc surveys or similar), it will be possible to have a more detailed analysis.

The approach taken considers as first “basic requisite for HNV forest areas” the exclusion of forests that are clearly of artificial origin (such as timber plantations) and exotic species stands.

The second step in the selection is a “high nature level” description related to the following main indicators:

- (a) *high nature value formations*: the nature and conservation interest of the formation, drawn from the list of natural and semi-natural habitats envisaged in EEC Directive 92/43 (“Habitat Directive”) and other considerations on the forest type;
- (b) *protected areas and stand structure*: presence of the stand in a protected area of European, national or regional interest and structure of the forest stand (old and mixed-age stands are considered important for biodiversity conservation).

The first part was based on the specific composition of forests as advised in the HNV areas interpretation manual. This analysis tries to assess the value of forest formations that correspond to the units of the Habitat Directive by using a “rarity” criterion as regards the representativeness of the inventory unit (inventory category or subcategory) at regional level.

To the formations identified with this procedure were added those deemed of conservation value anyway, even if they proved not to be rare in the administrative region considered (e.g. environments of special ecological interest for their biodiversity and the natural dynamics processes, formations of bio-geographical and landscape interest present in various regions).

The second indicator considers stands included in protected areas (national, regional or of European interest), for their intrinsic potential conservation value. Of these stands, only the ones that presented a forest evolution stage or a structure of nature interest, such as for example older-growth stages or irregular and uneven-aged structures, have been selected.

In a country like Italy that is characterised by a very extensive anthropic use of forests in past centuries, it seems reasonable to establish as “high value” features for forests, not only the presence of species or habitats of nature interest, but also elements of “conservative management” that safeguard natural development processes in forest ecosystems (and therefore also forms of human non-intervention). Thus, the indicators considered attempt to underline the importance of certain structural characteristics of the population that are fruit of a more or less active management (irregular and uneven-aged stands), but also of non-intervention (e.g., old-growth development stages of high forests and coppices), the latter being important for the natural evolution of the forest ecosystem natural dynamics.

Conclusions on land cover approach

More widely available data bases, such as CORINE and LPIS, may provide an approximate picture of the distribution of land cover types and patterns that are *potentially* HNV farmland, but generally will not

give an accurate calculation of the true extent. In particular these data bases might be used to show the approximate distribution of:

- ▶ Categories of pasture that are potentially HNV (CORINE, LPIS).
- ▶ Low-intensity crops in mosaics with a high proportion of semi-natural landscape elements (LPIS).
- ▶ Arable systems with a high proportion of fallow land (FSS, IACS data).

Where specialised inventories exist, land cover data may show the extent of some types of HNV farmland and forest, such as semi-natural grasslands, hay-meadows and orchards. Few countries have complete inventories of this sort for farmland. However, forest inventories are more widely developed.

In countries where suitable data exist, monitoring the extent of relevant land cover types and patterns provides valuable background information on tendencies in the physical components of HNV farming. Thus, if the extent of semi-natural woodlands, hay meadows, or traditional orchards is known to be declining, this is a direct indication of a specific HNV decline.

Further developing data bases and systems for monitoring the regional extent of such semi-natural land cover types is an important step in establishing effective monitoring of HNV farmland and forests.

The LPIS is an especially relevant and powerful tool where data could be incorporated on semi-natural land cover. Major progress in this direction has been taken in some countries, such as Bulgaria and Slovakia. These leads could usefully be followed by other Member States.

Overall, the land cover approach on its own has limitations, which need to be addressed through complementary methods (see below). Experience in several Member States suggests that it is not possible just with this approach to produce a baseline figure of the total extent of HNV farmland with sufficient reliability to be used as the basis for monitoring change. Rather, maps showing the distribution and approximate extent of different types of HNV farmland may provide a useful context for developing more specific indicators for these HNV types.

5.2.3.2 Farming systems approach

If available, data on farming practices should complement data on land cover, thus enabling some of the gaps discussed above to be filled. For example, an area of permanent grassland identified from land cover data may be semi-natural, or it may have been heavily fertilised. Farm systems data could clarify this by telling us the amount of nitrogen fertiliser applied per hectare on the farm in question. Alternatively, the livestock density per hectare will give a strong indication of the productivity of the grassland, and thus whether it has been reseeded and fertilised, or not.

In exceptional cases, relevant data can be found in national data bases. For example, the IACS data base in Austria records whether hay meadows have one, two or three cuts per year. This could provide an indication of intensity of use.

Generally though, most of the data needed are not available in current data bases. When they are available, the data usually are not geo-referenced at the level of the parcel or holding, and may be aggregated at the level of administrative regions. Data available as regional averages, for example of expenditure on agro-chemicals, are problematic as they may hide very large variations in practices at the

farm level. A region with a high average use of agro-chemicals may have some areas of highly intensive farming, and also significant areas of low-intensity HNV farming, but this would not be apparent from the data.

Data from the Farm Accountancy Data Network (FADN) may be relevant for some HNV criteria, such as input use. However, caution is required as economically small farms are excluded from the FADN surveys. Farms below the minimum economic-size threshold may include a high concentration of a region's HNV farming.

The following table summarises the farming practices most relevant to identifying HNV farming. Corresponding data should be sought at the holding or municipality level (not at higher levels of aggregation, unless reliable disaggregation is possible).

Examples of relevant farming practices	Possible data sources
Livestock densities per hectare of forage at the holding level.	Possible to extract data from IACS records, national agricultural census, etc.
Number of livestock that use off-farm seasonal grazings.	May be available from FSS, national agricultural census.
Number of cuts and timing of hay-making	Specialist inventories, sample surveys.
Number of livestock that are shepherded and/or practising transhumance.	Specialist inventories, sample surveys.
N use per hectare at the holding level.	Specialist inventories, sample surveys.
Biocide use per hectare at the holding level.	Specialist inventories, sample surveys.
Arable yields per hectare at the holding, or possibly municipality level.	Specialist inventories, sample surveys.
Management of arable fallow	Specialist inventories, sample surveys.
Management of understorey in permanent crops	Specialist inventories, sample surveys.

One type of data which may be widely available is livestock densities per hectare of forage. These data should be available from IACS records, and are highly relevant for identifying HNV farming systems.

In the absence of reliable statistics on the extent of semi-natural pasture and meadows, average livestock density per hectare gives a strong indication of the productivity of a given area of vegetation, and thus of whether the vegetation has been reseeded and fertilised. Thus a very low LU/ha of forage at the farm level indicates that the main forage used by that farm is semi-natural.

To be meaningful, such data should be calculated at the level of the farm holding. Average livestock densities for an administrative area or region are not a reliable indication of HNV livestock farming, as they may hide great variations in the density of livestock across the area. A possible exception would be in small areas (municipalities) where farms are known to operate a relatively homogenous system.

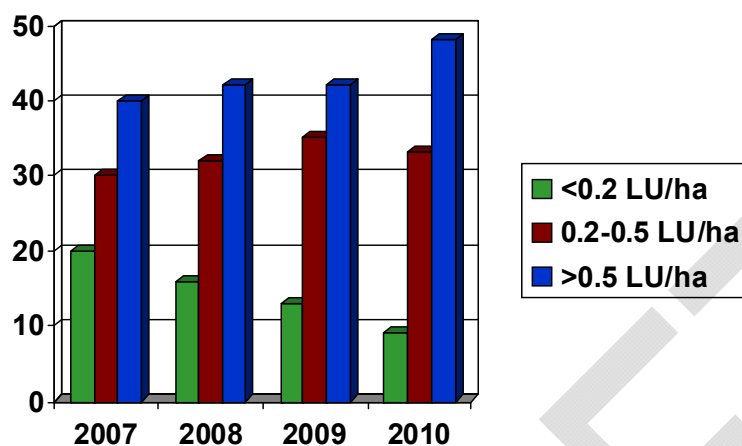
For a given region, HNV livestock farms will be found in the lowest bracket of livestock densities. In a region where grazing land is mostly of low productivity, HNV farming may be in a range of approximately 0.1 – 0.3LU/ha. The actual range that is used as an indicator must be determined according to regional and local conditions. Available research suggests that even in more productive regions, livestock farming based mainly on semi-natural vegetation is unlikely to exceed 1 LU/ha.⁸³

An estimate of the area of HNV livestock farming thus can be derived from the sum of forage land declared by farms in this "HNV stocking bracket". To be applicable for this purpose, stocking density

⁸³ See e.g. IEEP 2007

data must take account of all forage land, including off-farm grazing land, such as common grazings (note that these are not included in Farm Structures Survey – FSS). Vast areas of semi-natural grazing land fall into this category, and if excluded, highly distorted figures will be produced.

The following graph shows an imaginary monitoring of holdings in a region according to three categories of LU/ha of forage.



This approach is only possible and reliable where farmers' declarations of livestock numbers and forage area are an accurate reflection of the reality on the ground. For example, if farmers are able to claim SPS on the basis of a land area smaller than they are using in practice then there will be inaccuracies. Farms declaring common land may create problems due to double counting, or insufficient counting of land that is in current use.

5.2.3.3 Combining criteria to give an HNV "score" by farm or municipality.

HNV criteria may be combined in a points system to allocate an HNV "score" for a given unit of land, such as the farm holding or the municipality, on the basis of the characteristics that are present.

This is the approach followed in an analysis of HNV farmland in France by JRC. The JRC study (not implemented as an official indicator in France) uses a mix of data to allocate an HNV score to each municipality in France (NUTS5).

HNV points are allocated to each NUTS5 region according to a mix of variables, mostly land cover but also some farming system variables, using a variety of data sources:

- ▶ proportion of permanent grassland at farm level (FSS)
- ▶ diversity of crop types at farm level (FSS)
- ▶ area of common land
- ▶ proportion of land under "low-intensity" crops such as oats, rye and alfalfa, at farm level (FSS)
- ▶ presence of fish-ponds on farms (FSS)
- ▶ proportion of land under hedges and woodland edges (regional data)
- ▶ proportion of land under traditional orchards (regional data)

- ▶ use of N/ha of permanent grassland at farm level (national grassland survey)
- ▶ average crop yields (national agricultural statistics)

This HNV points approach applied at the farm level probably should include as a minimum the following variables:

- ▶ Proportion of the farm area under permanent grassland and/or arable fallow (available from IACS) – specific categories of low-intensity pasture more likely to be HNV may be selected, if available in data sets.
- ▶ Livestock density per ha of forage (available from IACS).
- ▶ Input use.

Additional variables to consider would be:

- ▶ Diversity of crops and other land cover types at farm level or within a given area (FSS or LPIS/IACS)
- ▶ Modal size of parcels at farm level or within a given area (LPIS)
- ▶ Proportion of land area under semi-natural landscape elements (hedges, ponds, patches) at farm or municipality level (potentially available from LPIS)
- ▶ Proportion of land area under common land at municipality level (national/regional data sources)

Data collected at farm level may be converted to averages at municipality level following JRC example for France, although caution is needed if municipalities are large and heterogeneous. The methodology is flexible and can be adapted to different characteristics of farming in different parts of Europe.

5.2.3.4 Species data

In most regions, data limitations mean that using species distribution alone is not a sound approach to identifying HNV farmland and forestry. The available data are not sufficiently exhaustive, either in terms of the range of species covered, or of geographical coverage, and are not up-dated with sufficient regularity.

On the other hand, species distribution data may be useful for filling in gaps left by the rather crude approaches currently possible using land cover and farming systems data. These are likely to miss some areas of farmland of high nature value, especially those characterised by smaller semi-natural elements and mosaic patterns, or farmland under more intensive use that continues to harbour species of conservation concern. Species data is particularly relevant for identifying Type 3 HNV farmland.

The JRC/EEA has applied this approach. Lists of indicator species for HNV farmland have been defined, and corresponding sites from Important Bird Areas and Prime Butterfly Areas have been selected and mapped, to complement the land cover picture.

Several countries have used national species data for this purpose in their HNV farming identification exercises, for example Bulgaria, Romania, Sweden and England.

However, this approach does not, on its own, make a link between biodiversity and farming systems or practices. Once areas have been identified on the basis of species data, it will be necessary to establish what land cover patterns and farming systems or practices are associated with the high level of biodiversity, in order to monitor changes in these characteristics.

5.2.3.5 Sample surveys

Current EU, national and regional data bases were not designed for monitoring tendencies in HNV farming. At best, they only provide time-series data on some of the HNV variables. Data collected at regional or national level therefore would need to be complemented by more detailed sample surveys.

Currently for most regions sample surveys are the only way to capture tendencies in certain variables, which may in fact be the most important ones, especially tendencies in:

- (1) The intensity of use on the farmed/forest area, and specific practices:
 - Meadows – reseeding, fertilisation, frequency and timing of mowing.
 - Pastures – fertilisation, grazing patterns, shepherding, transhumance
 - Permanent crops – management of understorey, fertiliser and pesticide use, maintenance of terraces
 - Arable – fertiliser and pesticide use, management of fallow
 - Forest – quantity of deadwood, harvesting systems, management of understorey
- (2) Condition of semi-natural land cover at the farm/forest level – semi-natural pastures, hay-meadows, orchards, woodlands, and smaller semi-natural landscape features such as hedges and ponds.

Example of sample survey approach to monitoring HNV farmland – Germany

Germany has taken the sampling approach to monitoring HNV farmland. A total of about 1,000 sites, each of 100 ha, is included in the survey. The sites were established originally for monitoring farmland bird species. Additional criteria, based on the HNV farming concept, have been incorporated. The system monitors the condition of relevant land cover elements, but does not monitor farming practices.

Under the applied method, an area or landscape element is classified as “agricultural land of high nature value” when its characteristics are of sufficiently high ecological quality. The assessment is on the basis of the diversity of botanical species.

Land units are allocated to five quality levels on the basis of a list of features, and are only assigned to the HNV farmland category once they have reached a certain minimum quality (Level 3 or higher). To this end, relevant assessment criteria have been drawn up for each land type or landscape element.

The listing of all units and landscape elements to be mapped and assessed, the assessment criteria and the additional mapping instructions are collated in a mapping manual and made available to the cartographers together with an aerial photograph for each sample area. The units are then surveyed on the ground using the technique of trans-sectional field walking. The mapping results entered on the aerial photograph are digitised and centrally collated. This ensures that the data are assessed uniformly at Federal level.

This method makes it possible to observe quantitative changes and also to record qualitative changes within the HNV farmland category. It is also possible to relate the development of HNV farmland to physical regions and regions which are defined according to ecological criteria (e.g. the North German Plain, the Alpine Foothills).

As the selection probability of the individual sample units is known, it is possible to extrapolate the overall quantity, i.e. the overall area of HNV farmland in Germany. Regular data gathering makes it possible to build up a picture of qualitative and quantitative changes in HNV farmland over time. This calculation is also possible for individual HNV farmland types (e.g. meadow orchards, HNV grassland etc.).

The statistically ingenious design of the survey minimises the cost of gathering data in the field and thus reduces the most significant cost factor. Coordination of data gathering across Germany, together with the use of a uniform

method, ensures the homogeneity of the gathered data at national level.

The simultaneous use of the survey method for different biodiversity relevant monitoring programmes (including farmland birds and HNV) opens up a number of possibilities for extended utilisation, so that the causes of any desired or harmful developments can be identified quickly, and appropriate management measures taken where required.

Some other countries, such as Sweden and UK, have landscape or countryside surveys that may be relevant for developing HNV survey methods.

Surveys of established sample sites could be complemented with random sample surveys outside these sites. Random sample surveys of farming practices are undertaken as part of FSS data gathering, and could be extended to cover HNV farming criteria.

5.2.3.6 Conclusions on quantifying the HNV baseline

It is important to keep in mind the purpose of measuring the extent of HNV farming and forestry, in the CMEF context. The aim is to be able to monitor changes in HNV farming and forestry systems, or in the main characteristics of these systems, compared with a baseline situation; and to assess to what extent these changes (or an absence of change) have been influenced by the RD programme.

Maps of HNV farming based on the JRC/EEA approach of land cover + species/habitats designated areas are not designed for monitoring tendencies in HNV farming at the level of a programme area. The boundaries of IBAs, PBAs and Natura 2000 sites are unlikely to be altered during the lifetime of an RD programme, so effective monitoring would need to focus on changes occurring to the HNV farming and forestry systems within sites.

In some countries with highly developed land cover and species data, effective mapping calculation of the extent of certain types HNV farmland and forest may be possible with this combination of national data. However, the slow pace at which it is possible to refresh the complex data involved means that it is unlikely that the impact of RDP measures could be reliably picked up by re-mapping at intervals during the programme. Also, this land cover + species approach will not shed light on changes taking place in HNV farming or forestry systems and practices, or how these have been affected by RD programmes.

Furthermore, any single programming region is likely to contain several types of HNV farming and forestry, each with particular characteristics. A single figure combining an estimated extent of all different types is not a sound basis for evaluating the effects of a programme.

For monitoring to be useful in practical terms (i.e. to inform the design of future RDPs), it is essential to gather information on the range of HNV farming and forestry systems, with their particular characteristics. It is important to know the trends in key land cover elements discussed in the land cover approach above, and also in the farming and forestry practices most relevant to HNV as discussed. This is a necessary basis for assessing the effects of RD programmes.

The precise method and combination of indicators will depend on the data, resources and preferences of each Member State. Approaches should be appropriate to the region. In some regions, HNV farming and forestry systems cover entire landscapes, and require quite complex sets of indicators for effective monitoring. At the other extreme are regions where HNV farmland and forest are limited to small areas that can be defined relatively clearly and monitored more simply.

5.2.3 Recommendable methods of measurement

In any given region, different approaches and methods for estimating the extent of HNV farmland and forestry are likely to produce significantly varying results. This reflects the reality of data sources that are far from perfect for the purpose. It is therefore advisable to try to capture a complete picture by approaching the question from a range of data sources and angles, and not to rely on a single approach.

Experience in several Member States suggests that it is not possible to produce a baseline figure of the total extent of HNV farmland with sufficient reliability to be used as the basis for monitoring change over the RDP period. Rather, the overall baseline extent can be considered as indicative.

Maps showing the distribution and approximate extent of different types of HNV farmland may then be combined with data on farming systems to provide the context for developing more specific indicators for each HNV type. The use of complementary sample surveys, as applied in Germany, is a highly recommended approach for monitoring these specific indicators.

Bringing together current practices and discussions at the Thematic Working Group meetings, the outcome would be a monitoring system at two levels (regional/national data + sample surveys) as follows:

Regional/national level – gathering available data on land cover and farming practices

Realistically, the best that can be achieved in most regions at present is to build an approximate picture of the extent of HNV farming and forestry in a region, or the extent of the most relevant characteristics, by drawing on the land cover and farming data that are available.

These may provide a basket of separate but complementary estimates of extent. Thus, land cover data will give one picture, and farming practices data will give another, complementary picture.

Data may be managed in statistical format, for example giving a figure for each indicator per municipality; they may also be translated into maps to help visualise the territorial distribution of HNV characteristics.

As a minimum, this regional data gathering should aim to provide indicators of:

- ▶ The extent of key HNV farmland/forest types as discussed above, if data on these are available.
- ▶ The extent of farmland/forest with key HNV farming/forestry system characteristics. Perhaps the most feasible at present is to monitor tendencies in the area under low-intensity livestock systems (LU/ha of forage at the holding level).
- ▶ The extent of farmland/forest with the presence of suites of indicator species that may be monitored to help with assessing the biodiversity condition of HNV land.

If databases allow, this regional/national picture may be developed through a method similar to that applied by JRC in its French study, allocating an overall HNV score by municipality, or holding. In this way, a weighting can be given to each characteristic.

These regional/national data sources, combined with expert knowledge, would allow the identification of broad, approximate zones with a high concentration of HNV farming and/or forestry systems, and the characterisation of systems within each zone. Precise delineation of such zones, and measurement of their extent, is not recommended, as this is not likely to produce a reliable indication of the baseline extent of HNV farming/forestry for monitoring purposes.

However, an approximate zoning exercise may provide a useful way of distinguishing broad HNV systems. These then provide the basis for discussion with multi-disciplinary experts in order to identify and get agreement on the key types of HNV farming and forestry in the region and the broad areas that they occupy.

For each HNV system, a set of indicators should be defined, preferably including at least one biological and one farming system indicator, to use in monitoring the impact of RDP measures on that system.

A sampling programme should monitor the baseline condition of each of these indicators and, as a minimum, repeat the sampling in the final year of the programme.

HNV sample surveys

Sample surveys should be designed to ensure full representation of the range of HNV farming and forestry systems in the programme area. It may be appropriate to design sample surveys that are specific to broad HNV zones, in order to select the most relevant characteristics that should be monitored in each zone.

The surveys should aim to monitor a range of HNV characteristics:

- ▶ Trends in key HNV farming/forestry practices, including input use, and practices such as shepherding, transhumance, management of arable fallows, late hay-cutting, management of understorey in permanent crops.
- ▶ Trends in the extent and condition of key types of semi-natural land cover, included small features such as hedges, patches and ponds.
- ▶ Species populations, covering a range of species associated with the local HNV farming/forestry system.

- ▶ Monitoring the socio-economic situation of HNV farming/forestry holdings is also extremely useful for subsequent assessment of the effects of RD programmes. However, this aspect is not a specific sub-indicator for HNV in terms of CMEF requirements.

Overall, these sample surveys will be essential for assessing the effects of RD programmes on HNV farming/forestry.

5.2.4 Data requirements and collection

Further development of existing data bases is an important consideration for the future of HNV farmland and forestry monitoring. It would be desirable to incorporate HNV variables in existing data bases, especially in FSS and LPIS-IACS, including:

- ▶ Parcels consisting of semi-natural farmland, including traditional orchards and hay-meadows, and smaller features such as hedges and ponds.
- ▶ Common grazing land used by the farm (area used in ha or LU grazing days).
- ▶ All forage land used by the farm (including scrubby and woody forage).
- ▶ All grazing livestock present on the farm.

First steps towards testing the incorporation of these data could be taken for the sample-survey sites, especially for LPIS.

5.2.5 Interpretation

CMEF Guidance Note A recommends the focus of evaluation should be on the bottom-up estimation of impact:

- ▶ *In a first step, impact should be estimated at the level of direct and indirect beneficiaries by programme evaluator on the basis of output and result indicators, survey data and benchmark data and coefficients from similar projects and past experience and evaluations (for calculation of double counting, deadweight, displacement and multiplier effects). This should be cross-checked against the counterfactual situation and contextual trends in programme area.*
- ▶ *In a second stage, the evaluator should make an estimation of the contribution to general trend at programme area level (baseline trend), where feasible/statistically significant compared to other factors. Where this is not possible the evaluator should make a qualitative assessment in general terms.*

The **Result** indicator for HNV is “Area under successful land management contributing to biodiversity and HNV farming/forestry”. It is only possible to say that a given piece of land is under “successful land management” if the land in question is under some sort of scheme that establishes the management that is required. In the case of natural forests, the management may be primarily non-intervention and protection.

Thus, “Area under successful land management contributing to biodiversity and HNV farming/forestry” will be composed of:

- ▶ Farms that have the range of HNV characteristics, as defined for the region and farming system in question.
- ▶ And that are participating in a scheme such as agri-environment, Natura 2000 compensation or LFA scheme, with management requirements that explicitly support the maintenance of the HNV characteristics and associated biodiversity.

The Result indicator therefore will give a quantified indication of the area of land under successful “HNV management”, which can be compared with the total estimated extent of HNV farming or forestry in the region.

Impacts of programmes should be evaluated in relation to the baseline indicators, both the qualitative and the quantitative indicators.

It is recommended that programme evaluators should treat HNV baseline figures in the RDPs, such as those based on a CORINE + designated areas approach, as a very provisional, indicative baseline. Any indication of quantitative changes, including no apparent change, must be interpreted with great care.

Indicators that combine several HNV variables to produce an overall HNV score or map may provide a broad background indication of trends. However, for the evaluation of impacts of individual measures, it will be more practical to use indicators that are kept separate from one another, and may be specific to particular systems.

For example, separate land cover indicators might show there has been a decline (or increase) in the regional extent of semi-natural hay-meadows, of traditional orchards, or of small-scale farmland mosaics. Farm systems indicators might show a change in the total extent of holdings within the HNV thresholds for LU/ha, or the number of livestock using seasonal grazings. Evaluators would make an assessment of the impact of individual RDP measures on each of these trends.

Simple numerical indicators cannot be devised that will indicate how rural development programmes are impacting on HNV farming and forestry. Rather, it is a question of using baskets of indicators to gather an understanding of how HNV farmland and systems are evolving, and then of using expert judgement to assess the role rural development measures may be playing in this evolution. Indicators may reveal conflicting trends, with some indicators pointing to a maintenance of the extent and condition of certain aspects of HNV farming and forestry, whereas others indicating a decline, or improvement in other aspects.

Programme evaluators will need to use their expert judgement and draw on all of the available information to make an informed assessment of the impact of the programme. The estimate of impact should reflect only that proportion of the change over time which may be attributed to the programme once the baseline trend and other factors have been taken into account. This requires an understanding of the causality between rural development interventions and any changes in HNV farmland/forest and systems characteristics, derived in part from a consideration of the counterfactual.

According to the CMEF guidelines, starting from the baseline analysis, authorities should *propose quantified objectives and targets (milestones where relevant, targets at the end of the period)*. The proposed

quantified objectives and targets for HNV farmland and forestry should provide a basis for the evaluation of individual measures and of the overall programme.

The Common Evaluation Questions in Guidance Note B do not refer specifically to HNV farming and forestry systems, although for Axis II there are several questions asking to what extent measures have contributed to the maintenance of “sustainable farming systems”, “the countryside”, “landscape and its features”, and “biodiversity”.

For all measures, it would be appropriate to consider the question: To what extent has the measure contributed to maintaining or improving the economic viability and/or ecological condition of HNV farming and/or forestry systems?

5.2.6 Judgment issues

The available quantitative indicators will not be sufficient to capture these complex relationships. The indicators are also likely to reflect changes in the environment arising from a variety of influences and decisions by different actors. The extent to which the changes observed can be attributable to rural development programmes will need to be inferred by programme evaluators on the basis of evidence available to them.

For example, one possible indicator for HNV arable systems in southern Member States might be an index of arable fallow land at farm or municipality level. Changes in the area of fallow land may be driven by a range of factors, including climate, commodity prices, rules attached to CAP Pillar I payments, and technological developments. The evaluation of the impact of RD measures would need to examine:

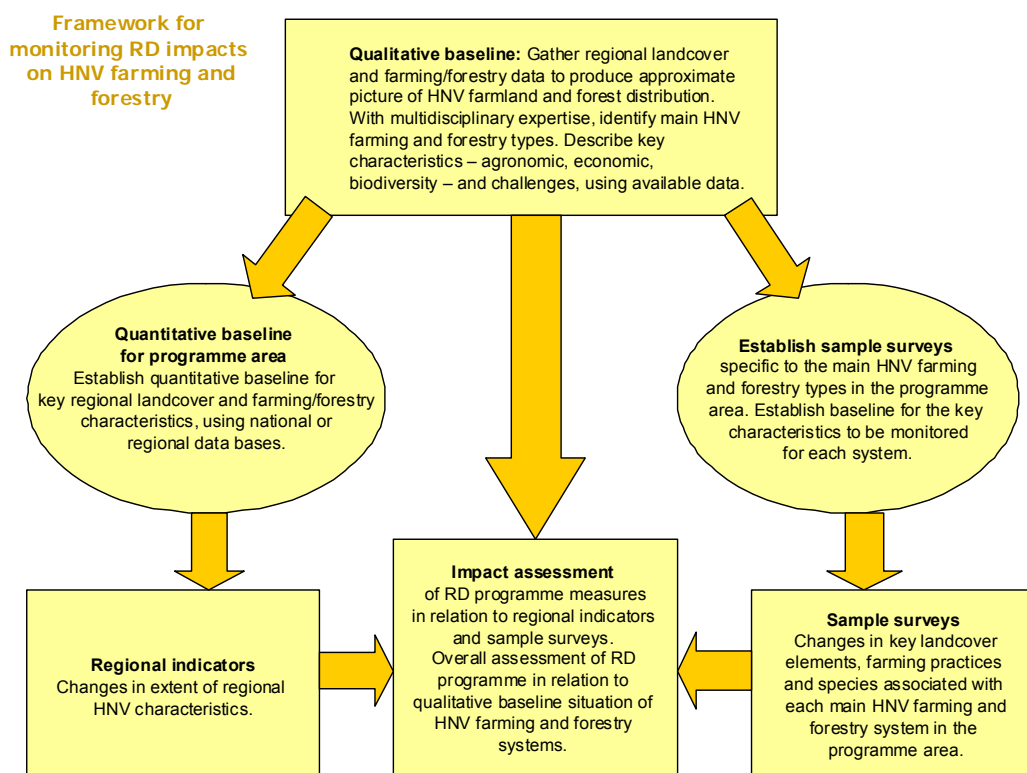
- ▶ Whether any of the RD measures includes mechanisms that are *intended* to influence farmers' decisions on fallowing. If so, what impact might be expected at the farm level.
- ▶ Measures that might *unintentionally* influence these decisions, and if so with what potential impacts at the farm level.
- ▶ How many farms and what extent of UAA have participated in the above measures.
- ▶ Whether the municipalities with the most significant changes in the area of fallow coincide with those that have participated in the above measures.
- ▶ Compare the relative potential impact of the measures with the perceived influence of other factors (climate, commodity prices, etc.).

Sample surveys of areas with a concentration of HNV farming and forestry systems will allow a far more rigorous assessment of programme impacts. The monitoring of individual farms, including farmer interviews, within the sample areas will make it possible to assess the impact of RD measures on the economic situation of HNV systems, on specific HNV farming practices, and on the ecological condition of farmland.

The inclusion of random sample surveys should allow a comparison of trends in HNV characteristics on farms that participate in RD measures, with trends on farms that do not participate.

Only an investment in appropriate data collection and monitoring schemes will ultimately allow a full evaluation of the effects of rural development programmes on HNV farming and forestry.

The following diagram provides an overview of a framework for monitoring HNV farming and forestry, to be used as basis for the assessment of impacts.



5.3 Impact Indicator 6: Improvement in Water Quality

5.3.1 The CMEF guidelines

This impact indicator is defined as (see guidance note J of CMEF handbook)

Changes in gross nutrient balance (GNB)

Quantitative change in the estimations of GNB that can be attributed to the intervention once double counting, deadweight, and displacement effects have been taken into account. The GNB indicates potential nutrient losses to the water bodies likely to be detrimental for the quality of water.

The data collection on the different programming levels is supposed to be conducted:

- (1) by estimation by programme evaluator at the level of direct and indirect beneficiaries on the basis of output and result data, survey data and benchmark data and coefficients from similar projects and past evaluations and modelling work (for calculation of double counting, deadweight, displacement).
- (2) by cross-checking against counterfactual situation and contextual trends in programme area, particular as regards relevant driving forces, pressures and responses.
- (3) by estimation of the contribution to the general trend at programme area level (baseline trend), where feasible/statistically significant compared to other factors.

Related Common Evaluation Questions and Horizontal Evaluation Questions according to Guidance Note B

Measure Code 213

Evaluation questions: To what extent have compensatory allowances contributed to effective land management in river basin areas affected by the WFD? To what extent have compensatory allowances contributed to maintaining the countryside and improving the environment?

Measure Code 214

Evaluation questions: To what extent have agri-environmental measures contributed to maintaining or improving water quality? To what extent have agri-environmental measures contributed to maintaining or improving soil quality?

Measure Code 215

Evaluation questions: To what extent have the payments contributed to encouraging farmers to adopt high standards of animal husbandry which go beyond the relevant mandatory standards?

Measure Code 216

Evaluation questions: To what extent have supported investments contributed to maintaining the countryside and improving the environment?

Measure Code 225

Evaluation questions: To what extent have forest-environment payments contributed to maintaining or improving water quality? To what extent have forest-environment payments contributed to preventing soil erosion?

Measure Code 226

Evaluation questions: To what extent have the supported actions contributed to improving the environment?

Measure Code 227

Evaluation questions: To what extent have the supported actions contributed to improving the environment and maintaining the countryside?

5.3.2 Key challenges with regard to measurement and interpretation

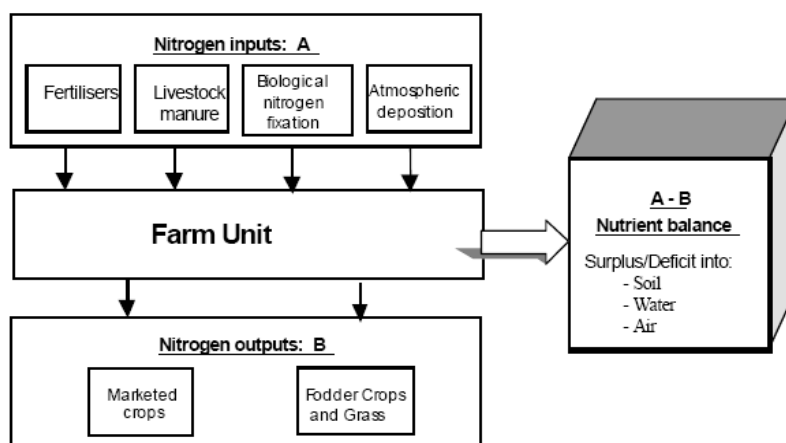
Natural site conditions determine the impacts of agricultural production on water quality directly and indirectly via the farm management. Regarding direct effects natural conditions such as slopes, denitrification in the soil and ground water body as well as water residence times play an important role in regard to nutrient leaching. Indirect effects are consequences from the farm management that includes the application of RD-measures. Against this background the key challenges with regard to the measurement of the impact indicator and its interpretation regarding water quality are as follows:

- ▶ Separate the effect of RD-measures that are applied under alternative combinations from conditions on the change of GNB. Due to interdependencies between RD-measures, the total impact (expressed in changes of GNB) cannot be simply calculated as the sum of impacts of single measures. Issues of double counting, deadweight, and displacement play an important role since there are several other factors which might influence the GNB.
- ▶ The GNB indicates the amount of nutrients that can be potentially emitted into the water and should be interpreted as a potential risk indicator for water quality only (see 5.3.5). Indeed, there are many variables influencing the transfer of nutrients from the soil to the water bodies to establish a direct and simple relationship between GNB and nitrogen concentration in water.

5.3.3 Recommendable methods of measurement

Several methods have been developed for assessing agricultural nutrient balances. CMEF uses the gross nutrient balance, a method developed and recommended by the OECD. A full explanation of the gross soil surface balance is provided by the OECD/Eurostat Nitrogen Handbooks (OECD/Eurostat, 2003). In principle, the GNB includes all residual nutrient emissions of environmentally harmful compounds from agriculture. A detailed description of the OECD gross nitrogen balance and application for Germany is given in Panten et al. (2009).

Figure 1 The main elements of the nitrogen balance calculation



Source OECD 2007 www.oecd.org/tad/env/indicators

The soil surface balance calculates the difference between the total quantity of nitrogen inputs entering the soil and the quantity of nitrogen outputs leaving the soil annually. The N surplus results are expressed in kg of N ha⁻¹ of agricultural land.

The annual total quantity of nitrogen inputs includes:

- (1) Inorganic or chemical nitrogen fertilizer: quantity consumed by agriculture
- (2) Net livestock manure nitrogen production: total number of live animals in terms of different species, sex, age and purpose, multiplied by respective coefficients of the quantity of nitrogen contained in manure/animal/year
- (3) Atmospheric deposition of nitrogen: total agricultural land area multiplied by a single coefficient of nitrogen deposited/kg/ha
- (4) Biological nitrogen fixation: area of harvested legume crops multiplied by respective coefficient of nitrogen fixation/kg/ha
- (5) Nitrogen from recycled organic matter: quantity of sewage sludge applied to agricultural land multiplied by a single coefficient of nitrogen content/kg of sewage sludge

The annual total quantity of nitrogen outputs or nitrogen uptake includes:

- (1) Harvest crops: quantity of harvested crop production multiplied by respective coefficient of nitrogen concentration
- (2) Forage crops: quantity of forage crop production multiplied by respective coefficient of nitrogen concentration

The soil N surplus is therefore estimated as being the difference between N inputs and outputs.

In view of the multitude of variables that determine the level of GNB, any changes in the CMEF impact indicator for water quality can be caused by various factors inter alia by RD measures. As mentioned above the key challenge is to separate the effect of RD measures on the change of GNB from other factors. Examples for these other factors are a change in mineral fertilizer use, or a change in dairy cattle raising.

The GNB can be calculated for a variety of spatial scales if adequate data is available. However, the interpretation and significance of the GNB and its changes in regard to water quality is different since several natural conditions and processes not measured determine the amount of nutrients leaching into the water (see 3.6.5).

Micro level

The farm is the management unit of the agricultural system and therefore represents the unitary micro unit. Due to the key challenges mentioned in paragraph 5.3.2 the farm is the appropriate level to measure the impacts of RD-measures on changes of the GNB. Several methods have been developed for assessing a farm nutrient budget. They are based either on a synthesis from individual fields (Benoit 1992) or on an analysis of the farm as a whole (Schroder et al. 2003). The later is more realistic since it also takes into accounts transfer of matter between fields and farming practices.

The most appropriate method to determine empirically the impact of RD measures and combination of RD measures on the change in GNB is the “difference in difference (DiD)” approach (see Chapter 3 Evaluation of Impacts):

- ▶ The first step is to select a group of beneficiaries and a control group of non- beneficiaries. The selection is crucial since the level of GNB depends not only on various natural site conditions such as soil and climate but also on specific farm characteristics such as structure and specialisation of production.

Example from Italy

- (1) Definition of the unit of reference (UR), in Italy the cadastral sheet (a polygon that is an average of 100 hectares) is used, although for some information more aggregate data (municipality NUTS5 or province NUTS4 may be used);
- (2) Classification of the territory according to agro-pedology and environmental characteristics (climate, geomorphology, soil, irrigated areas);
- (3) Quantification of the agricultural units for crop type, action and for UR;
- (4) Quantification of current total agricultural units that include both those conducted with conventional techniques and those following various actions (organic farming, integrated production, etc.);
- (5) Definition of technical itineraries for single crop and for the homogeneous areas identified in step 2, both in conventional farms that in farms benefiting of measures. Inputs of mineral N+ manure and their mode of administration, use of pesticides – active periods and doses, grassing for arboriculture, irrigation. This data can come from farm surveys or from interviews with experts (field agronomists, farmers, local technicians, etc.);
- (6) Estimation of total inputs of nitrogen and phosphorus (mineral + organic) and pesticides on farms receiving agri-environmental payments and on conventional farms;
- (7) Estimation of surplus of nitrogen and phosphorus by nitrogen and phosphorus balance at farm level;
- (8) Estimation of input and total surplus of nitrogen and phosphorus and pesticides in the investigated territory or farmland level through the results of Sections 3, 4, 5, 6 e 7;
- (9) Quantification of indicators for estimating the effects of environmental measures on the input and on the surplus:
 - Variation in total input of nitrogen, phosphorus and pesticides (kg and %) induced by any single measure;
 - Variation in surplus of nitrogen and phosphorus (kg and %) induced by any single measure.

Challenges/limitations

- (1) Difficulty in knowing the levels of use of minerals and organic fertilizers (input) for each farm: asking this information directly from the farmer has not always been appropriate, as farmers are reluctant to give too detailed information on their farm practices. However, well-designed questionnaires can succeed in obtaining more reliable information;
- (2) The transition from micro to macro level is very complicated, due to a lack of information from non-beneficiaries (unrepresentative sample). In the latter case context data should be used but it often does not have the proper territorial (i.e. scale) level or is not updated. There is the need to work on a smaller scale of detail than NUTS4, because the regional farmland is very different with regard to landscape, agricultural practices, morphology, soil, climate, etc.;
- (3) There is an agreement to share the methodology, for example using the same coefficients for the calculation of

the surplus of nitrogen; but it is by far more difficult to harmonize the basic data for the analysis. In Italy there are regions with a high level detailed data available while in other regions data is practically absent.

- ▶ In the second step gross soil surface nutrient balance are calculated using the concept described above for beneficiaries and non-beneficiaries. The impact of RD measures on the change in GNB results from a comparison of the development of GNB between both groups.
- ▶ A third step is to aggregate the impact from single measures and measure combinations at farm level to regional, national and program level. The aggregation method should avoid double counting, deadweight, and displacement effects.

Example from Poland

In Poland the nitrogen balance calculated following the OECD methodology at the field level. The calculation of balance takes into account the loads of N associated with manures and mineral fertilizers as well as an uptake by crops grown. The amounts of manure N were assessed from number of stocking density reported for NUTS-5 level whereas the mineral N was desegregated from a regional level. The structure of crops grown was taken from paying agency databases (ARiMR) and aggregated to a village level (ca. 150 ha in size).

The assessment of RPD 2004 impact on water quality done based on N balance was verified by comparing its results with data from ground water monitoring system established for farmland in Poland (Igras 2008). Results are presented as average differences for different RDP instruments compared to reference areas not participating in RDP. Evaluation is based on 4000 geo-referenced monitoring profiles which allows for a reliable assessment of different instruments impact on the risk for N transport.

Since the approach requires detailed farm management data it might not be feasible throughout the Member States.

Macro level

The macro level is the farming region or farming territory. It is identified as the geographic entity with similar geology, soil and climate and the social groups which occupy it and interact there (Papy 2001). Therefore, the geographical limits of a farming region can be extremely variable. If the farm is the unit, the farming region is the appropriate scale to evaluate the interactions between farms which constitute an emerging property of agriculture at this organisation scale (Cristofini 1985). As a consequence, at the farming region level the environmental impact, including the soil N surplus cannot be considered as the simple sum of the impacts of each farm.

A wide range of methods has been developed for assessing the environmental impact of a farming region. Payraudeau & van der Werf (2005) classify these methods into 6 categories (Table 1).

Table 3 Six methods for regional environmental impact assessment (from Payraudeau & van der Werf 2005)

Method	Case studies author(s)	Object and scope
ERM: environmental risk mapping	ERM-1: de Koning et al. (1997) ERM-2: Giupponi et al. (1999)	Modelling of the soil nutrient balance as a sustainability indicator: nation scale (Ecuador) Modelling impacts on water quality of alternative land use scenarios with the GLEAMS model and calculation of environmental impact indices: lagoon of Venice catchment (Italy)
LCA: life cycle analysis	LCA-1: Biewinga and van der Bijl (1996) LCA-2: Geier and Köpke (1998)	Evaluation of ecological and economic sustainability of energy crops using a range of indicators, mainly based on LCA: Europe Evaluation of a complete conversion from conventional to organic farming using LCA: extrapolation at the rural area scale (Germany)
EIA: environmental impact assessment	EIA: Rodrigues et al. (2003)	Evaluation of the sustainability of agricultural technology innovation by the EIA method: Field-Farm (Brazil)
MAS: multi-agent system	MAS-1: Petit et al. (2001) MAS-2: Becu et al. (2004)	Evaluation of the quantity and quality of groundwater by using hydrological, agronomic and socio-economic models in a multi-agent system: Beauce aquifer (France) Modelling the impact of catchment irrigation management under social and agronomic constraints in a multi-agent system: catchment scale (Thailand)
LP: linear programming	LP-1: Zander and Kächele (1999) LP-2: Hengsdijk and van Ittersum (2003)	Optimisation of different production systems described at the farm level with multiple goal linear programming: extrapolation to a regional scale (Germany) Optimisation of different production systems to maximise the production targets whilst minimising impacts: farm or regional scale (Mali)
AEI: agro-environmental indicators	AEI-1: ECNC (2000) AEI-2: Rasul and Thapa (2004)	Development of "Driving force", "State", and "Response" indicators for environmental impact assessment mainly at macro-level: European agriculture Evaluation of farming sustainability by ecological, economic and social indicators: micro-regions (Bangladesh)

Among the 11 existing methods of regional environmental impact assessment, only 6 take soil quality specifically into account, i.e. ERM-1; EIA; LP-1; LP-2; AEI-1; AEI-2 (see reference in the Table above).

Several models have been developed recently to estimate soil gross or net nutrient balance at the NUTS2 or NUTS3 region levels in Europe. In contrast to farm level approaches that face the problem of aggregating GNB to the national level the challenge here is to regionalize the data that is only available at sector level e.g. total purchases of mineral fertilizer use from the national agricultural accounts. Such approaches are for example implemented in the agricultural economic sector models such as CAPRI (www.capri-model.org) for the EU-27 NUTS2 regions or in RAUMIS (Heinrichsmeyer et al. 1996) for the NUTS3 regions of Germany. The advantage of these models is the coherent modelling framework that allows inter alia for corrections of implausible statistical data. A central problem of top down approaches is the fact that the developments of important variables such as mineral fertiliser demand reflect the impacts of all influencing factors. Hence, an empirical quantification of impacts of RD measures and measure combinations on GNB would be far from significant. In general, these impacts are represented by fixed coefficients that are taken from the literature for relevant farm management practices that are supported by RD measures.

5.3.4 Data requirements and collection

Micro level

The mass balance is rather difficult to calculate as it requires the real N input into the farm. In order to apply the suggested "difference in difference" approach detailed information on nutrient amounts used on farms are required. In Germany, for example, such data should in principle be available on farms since farmers are supposed to calculate so called annual and multiyear nutrient management plans to record the farm nutrient in- and outputs. However, while beneficiaries of RD measures should provide the necessary data for this exercise, this cannot be expected from non-beneficiaries.

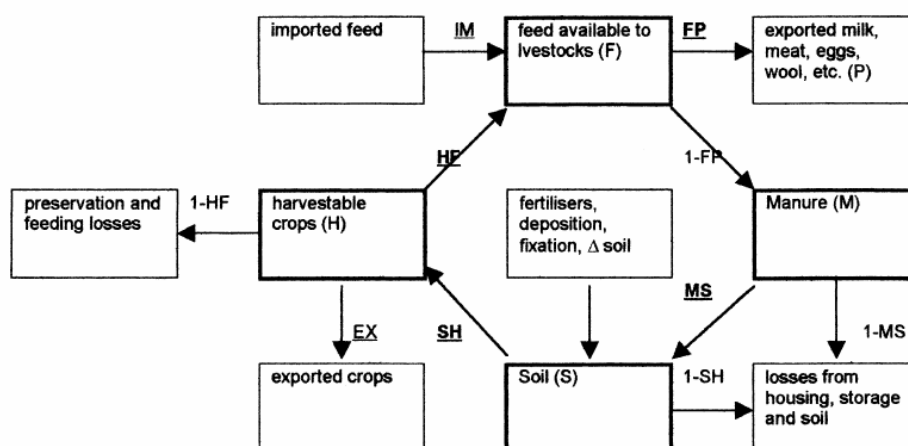
Macro level

At the farming region level the percentage of different land cover, land use and farming types are the key data required to evaluate the gross N balance together with atmospheric N fixation and deposition. For the quantification of land use and land cover a variety of data sources are available. Besides the farm structure survey (FSS) and ICAS data land cover can be determined due to the development of remote sensing devices, digital elevation models and GIS software. All European countries have access to these tools and are currently using them. Now about 20 years of data sets are existing, which allow analysing land use and cover changes with high spatio-temporal accuracy. Data on livestock is provided also by the FSS and livestock surveys.

Yet, the main issue with these models is to translate land cover into land use (e.g. from grassland land cover to fertilized, heavily grazed or hay production land use) in order to apply the most relevant coefficient of N input and output (Verburg et al. 2009) and to harmonise them within the European Union. This will not solve the problem of the lack of evaluation of the results obtained (see 5.3.5) but at least it will allow comparison of the results obtained.

The GNB represents a simplification of complex interrelated and variable processes. The calculation of GNB comprises uncertainties that are mainly based on the statistical data base or the coefficients used to convert statistical data into nutrient quantities e.g. livestock into organic nutrients. Specifically, if the amount of mineral fertilizers can be reasonably determined, the N application in manure is not obvious since it depends on the type of manure, the storage method and retention time before application. The second problem is related to the fact that N balance is the result of transfer of N mass from one element of the dynamic farming system (e.g. fertiliser, manure...) to the other (e.g. crop, crop residue, fodder, animal production ...). This transfer of mass is accounted for by using conversion coefficients (Figure 2). Schroder et al. (2003) identify 4 main conversion coefficients: SH to convert soil-N into harvestable crop-N; HF to convert harvestable crop-N into feed-N; FP to convert feed-N into animal produce-N and MS to limit the gaseous losses from stables, manure storages and manure application (Figure 2).

Figure 2 Flow diagram for a mixed farming system with conversion coefficients SH, HF, FP, MS and out-boarding fractions IM, EX



from Schroder et al. 2003

The main challenge for the future is to harmonise these conversion coefficient within the EU Member States in order to compare the results. It entails the publication of very different results from different countries or regions. For instance van Eerd & Fong (1998) found that nitrogen manure coefficients from northern European countries varied by 80% for heifer, 100% for dairy cow, and 400% for pig. This harmonisation is in particular essential for comparing GNB levels. However, the issue is of less importance for the comparison of GNB-changes as required by CMEF.

5.3.5 Interpretation

The gross nitrogen balance represents the theoretical nitrogen surplus in the soil calculated by the difference between the total quantity of nitrogen inputs entering the soil and the quantity of nitrogen outputs leaving the soil annually. **The use of gross nitrogen balance as an indicator of the potential N loss to aquatic system is significant.** Gross nitrogen balance does not inform on the form (organic, ammonia, nitrate) in which nitrogen is in the soil. If nitrate is the form much more prone to leaching, organic N is rather stable and is function of the carbon concentration in the soil. A better evaluation of the N risk to water quality would require estimation/measurement of gas emission (Net Nitrogen Balance). Ideally, water quality monitoring (nitrogen fluxes measurements at the outlet of agricultural catchments) would be the best method.

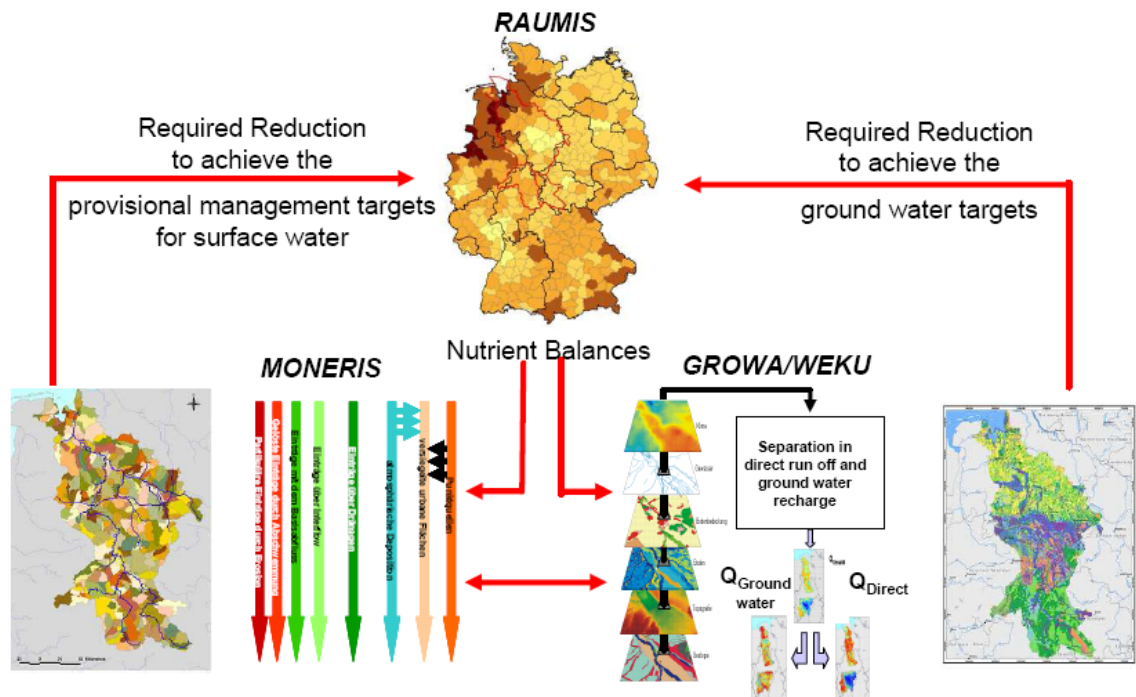
Aggregating farming region gross nitrogen balances at the country scale does pose specific limitations in information because it masks the heterogeneous reality of responses which are expected from different regions due to inherent intrinsic differences such as geology, soil, climate and socio-economic context. Aggregating scores by country, if necessary, should be a function of the rate of change of local/regional agricultural areas. Therefore, the results should be expressed as a rate of change of estimated N surplus by “representative region” (NUTS3 or smaller). This approach would provide several useful indicators to quantify the impact of rural policy on water quality. It would allow to: i) measure the impact of rural policy on water quality in different agricultural contexts; ii) determine the most “receptive” contexts where improvement is noticed and the least “receptive”, iii) analyse the reasons of such differences at the appropriate scale, iv) compare the rate of change in “similar” agricultural regions among countries.

The long history of catchment research has demonstrated some reasonable correlations between the proportions of land-cover types and nutrient export from drainage basins and in turn water quality. The method, with many more or less sophisticated derivatives, attributes a potential leakage factor for nitrogen to different types of land cover (crop, meadow, pasture, forest...) found in the drainage basin. Modelling runoff and water quality involves some degree of spatial averaging (or ‘lumping’) of land covers within the catchment. In several models the drainage basin is divided into several subcatchment units, each of which is treated as a uniform unit. Some studies of larger river basins have even adopted a fully lumped approach. Therefore impact of a given policy on water quality is evaluated indirectly via the change in land cover within a given drainage basin. This change in land cover is then simply translated in potential nutrient fluxes at the outlet of the catchment by using coefficient factors.

Recently, model networks have been developed to link regional gross soil surface nutrient balances to water quality. They represent approaches to overcome the lack of interpretation of GNB on water quality which is the second challenge (see 5.3.2). The model networks are very useful to identify so called hot-spots and may help regionally target RD measures as mentioned.

For instance Leip et al. (2007) developed a modelling framework that links the large-scale economic model for agriculture CAPRI with the bio-geochemistry model DNDC to simulate greenhouse gas fluxes, carbon stock changes and the nitrogen budget of agricultural soils in Europe. Kreins et al. (2009) developed an interdisciplinary model network consisting of the regionalized agricultural and environmental information system RAUMIS (Heinrichsmeyer et al. 1996), the hydrogeological model GROWA/WEKU (Wendland et al. 2002; 2004) and nutrient emissions in river systems MONERIS (Behrendt et al. 1999) to analyze the impacts of nutrient reduction measures on the water quality of a 49,000 km² catchment in Germany. Within this modelling network RAUMIS calculates the regional soil surface gross nitrogen balance (Figure 3).

Figure 8 Interactions of models for the Weser River basin.



from Kreins et al. 2009

The model network was used to detect “hot spot” areas of diffuse agricultural water pollution at the drainage basin level. The “Hot spots” represent target areas where rural development e.g. agri-environmental measures can help achieve the WFD objectives.

5.3.6 Judgment issues

The gap between gross nitrogen balance measurement and the impact on water and air quality could be easily filled, i) by measuring water quality in agricultural catchment and ii) by using net nitrogen balance which implies to estimate N gas emissions. This would allow developing a real systemic appraisal of the consequences of the agricultural policies and practices.

Water quality measurement

First, it has to be emphasised that the assessment of impacts of RD measures in the field of water quality by using GNB shall provide foremost information on nitrogen leaching/evaporation risk but not of direct water/air pollution. It is therefore a tool to evaluate the potential impact of RD measures on the environment – in particular water. It should show whether RD measures increase or decrease the risk of water/air pollution. – Thus when assessing this indicator, the currently available information (e.g. from WFD) shall be adapted to the necessities for capturing the impacts.

The measurement of water quality parameters, including all the nitrogen species in streams, lakes and groundwater is the most accurate method to assess the water quality and its evolution. This is the basis of the EU Water Framework Directive (WFD). There exists already an important network of monitoring sites throughout Europe with rather long-term data sets (usually up to 30 years of data). The primary goal of this WFD monitoring is to assess the water quality of streams, lakes and groundwater and follow its evolution. The data set gathered is compared against water quality thresholds which allow classifying water bodies.

Example from Estonia

There are 3 water quality indicators in the RDP of Estonia.

1. Quality of the drainage water

Drained fields with different environment support schemes are selected for the monitoring of drainage water quality in the different places of Estonia. Water samples from the drainage system are collected two times and the discharge is measured four times in the month. The concentration of NO₃⁻, NH₄⁺, P, K and S of water samples are measured in the laboratory. On the base of plant nutrient concentration in the drainage water and the discharge, the leaching of every plant nutrient is calculated (kg/ha/year).

On the same test fields the total nutrient balance is calculated to compare the influence of different environment support schemes on the usage of fertilizers.

2. Gross nutrient balance

Data is collected annually from the year 2004 from approximately 120 agricultural enterprises. Monitoring companies and farmers are chosen from different Agri Environment Support (AES) scheme types: supporting Organic Farming (OF), Environmentally Friendly Production Scheme (EPS) and Environmentally Friendly Management Scheme (EMS). In addition to support scheme types companies are assorted by type of farming (field crops, livestock or mixed (plant production + livestock)) and by size (<40 Ha, 40-100 Ha, >100 Ha).

Total balance of nutritional cells is elaborated on farm level and calculated 01.January – 31.December (corresponding also to the economical accounting period). Data is collected from farmer's field book, from their accountant's data, and by surveying enterprises. Data is collected, stored, and processed electronically.

Nutritional cells' input (bought into farm) and output (sold from farm) is compared in nutritional cells balance on a certain period of time according to one farming company, this enables to evaluate (through nutritional cells balance) the effective use of nutritional cells, surplus and shortage (kg/HA). Inputs (re-calculated into nutritional cells N, P, K g per kg dry matter) bought into farm: fodder, straw, mineral fertilizers, seeds, livestock, organic manure, nitrogen fixation by legumes, rainfall deposition etc. Outputs (re-calculated into nutritional cells N, P, K g per kg dry

matter) sold from farm: livestock, manure, feed and straw, crops etc. Unsold input in the end of manufacturing year

carries over to the next calculating year.

3. Pesticide use

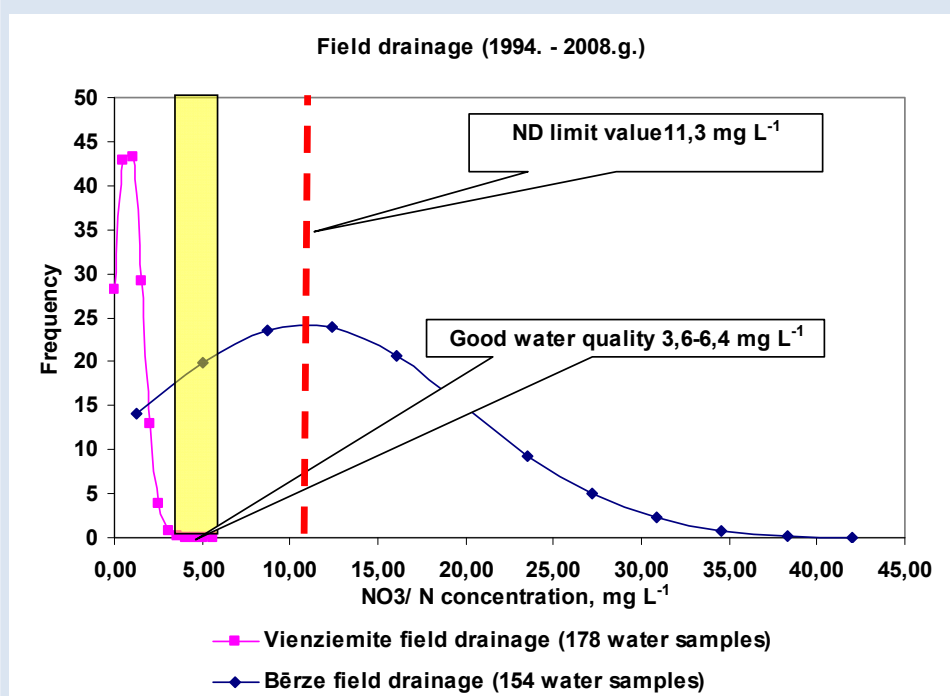
During pesticide use study, data is collected from the Environmentally Friendly Production (EPS) and Environmentally Friendly Management (EMS) manufacturers, who are given the benefit. Pesticides use and number of pesticide treatments are also elaborated on farm level and calculated 01.January – 31.December (corresponding also to the economical accounting period). Following data is collected during monitoring:

Cultivated cereals – type, treated and non-treated ha, yield (t/ha); Pesticide treatments –number of treatments, cereal type, ha, seed dressing (name, dose), herbicide (name, dose), fungicide (name, dose), insecticide (name, dose).

Example from Latvia

Agricultural run-off monitoring data collected since 1994 in small catchments' and drainage field scale has been used. Reference values for nutrients concentration (mg L^{-1}) for water quality classification (5 classes according WFD) could be designated using following percentiles: (1) <10% excellent quality, (2) 10 – 25% good quality, (3) 25 – 75% fair quality; (4) 75-90% poor quality; (5).> 90% bad quality. Figure below shows the distribution curves for drainage field Vienziemite with low input agriculture and drainage field Berze with intensive agriculture.

Probability (normal distribution) curves for the nitrate values evaluation in the small catchments' run- off



Nutrient criteria numeric ranges, developed at the national level using existing databases and agricultural run-off monitoring, could be used to derive specific criterion values of water quality used to evaluate the success of RD plan.

The most important finding of the research is that water quality standards for drainage water as well as for small catchments with intensive agriculture should be less stringent than for rivers, otherwise it will not be possible to fulfill the objectives – good water quality status in year 2013, set by the WFD.

Recommended classification for values of the nutrient quality status could be used for the purpose of rural development monitoring and evaluation, e.g. field drainage water quality could be used to assess the impact of soil surface GNB.

Net nitrogen balance

The intrinsic problem of relating the change of gross N balance to a change in water quality is that this N surplus comprises in fact 3 different N components, i.e. the soil N stock, the N leaching to water

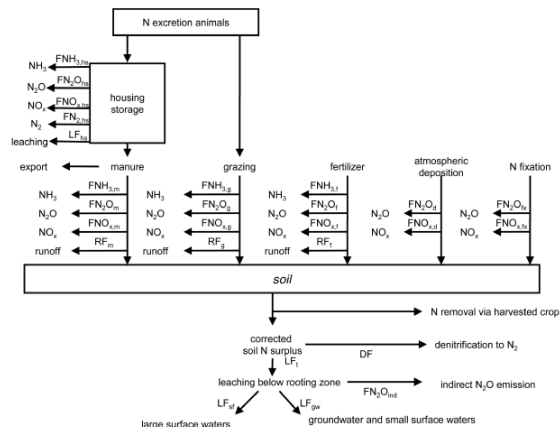
bodies and the N emitted into the atmosphere. The proportion of each cannot be easily assessed. Yet, this is fundamental to know if the surplus N is stored in soil and could be used for further production, or if it is transferred to water bodies and therefore influences the water quality, or if it is emitted in the atmosphere as inert gas (N_2), potent greenhouse gas (NO , N_2O) or as NH_3 to contribute to soil and water acidification.

Combining N surplus calculation with real water quality data on nutrient fluxes (see section “Water quality measurement” above) could help to reduce these uncertainties in future. It would require an adjustment of the sampling strategy of the WFD. It would provide true values for N leaching in agricultural areas and reduce the N surplus uncertainty by analysing the real consequence of rural policy on water quality.

The use of existing models estimating the gas emissions would reduce the uncertainty of the soil nitrogen balance and allow evaluating the consequences of a given policy or practicing on greenhouse gas emissions. Moreover, it could be used for the Contribution to Combating Climate Change indicator. This is getting urgent since the recent study from Schulze et al. (2009) emphasises the fact that the high emission of greenhouse gas (CH_4 & N_2O) in European landscapes offsets their potential capacity to uptake and store carbon,

For instance, MITERRA-EUROPE (Figure 9) partly based on the existing models CAPRI (<http://www.capri-model.org>) and GAINS (<http://www.iiasa.ac.at/rains/gains>) allows estimating nitrogen gas emissions (Velthof et al. 2009).

Figure 9 Schematic presentation of MITERRA-EUROPE



5.4 Impact Indicator 7: Contribution to Combating Climate Change

5.4.1 The CMEF guidelines

Definition of indicator according to Guidance Note J

The common impact indicator

The CMEF common impact indicator for climate change is defined in Guidance Note J as the “Increase in production of renewable energy”, measured in units of ktoe (kilotonnes of oil equivalent). The indicator is defined as “*quantitative and qualitative change in the production of renewable energy that can be attributed to the intervention once double counting, deadweight, and displacement effects have been taken into account*”. The assessment shall be conducted by the estimation by the programme evaluator at the level of direct and indirect beneficiaries on the basis of output and result data, survey data and benchmark data and coefficients from similar projects and past evaluations and modelling work (for calculation of double counting, deadweight, displacement). It should then be cross-checked against counterfactual situations and contextual trends in the programme area (i.e. useable agricultural areas and investment in renewable energy plant). Based upon this an estimation of contribution to general trend at programme area level (baseline trend) should be conducted, where feasible/statistically significant compared to other factors. The indicator is meant to be reported on *ex ante*, mid-term (2010) and *ex post* (2015).

Additionally – in order to arrive at a comprehensive assessment of the impact of RD Programmes in the field of climate change the baseline indicators as well as the common result indicators may be taken into account:

Relevant common baseline indicators

24 – Climate change: Production of renewable energy from agriculture and forestry

This indicator is defined as Production of renewable energy from agriculture and forestry and measured in Renewable energy from agriculture: KToe (1000 tons of oil equivalent), Renewable energy from forestry: KToe (1000 tons of oil equivalent).

For this indicator, due to data availability, production of renewable energy from agriculture covers biofuel:

- ▶ Biodiesel from oilseeds crops
- ▶ Ethanol from starch/sugar crops

Therefore it does not cover:

- ▶ Energy from short rotation forestry (on agricultural land)
- ▶ Energy from agricultural biogas (livestock manure)
- ▶ Energy from cereal straw

Production of renewable energy from forestry covers:

- ▶ Purpose-grown energy crops (poplar, willow, etc.)
- ▶ Woody material generated by an industrial process (wood/paper industry in particular) or provided directly by forestry and agriculture (firewood, wood chips, bark, sawdust, shavings, chips, black liquor etc.)
- ▶ Wastes such as straw, rice husks, nut shells, poultry litter, crushed grape dregs etc.

Conversion coefficients from ktors to ktce (EEA – IRENA 27):

Bioethanol: 0.800 kg/L – 23.4 MJ/L – 41,868 ktce/GJ

Biodiesel: 0.875 kg/L – 33.0 MJ/L – 41,868 ktce/GJ

25 – *Climate change: UAA devoted to renewable energy*

The indicator is measured as hectares (Ha) of UAA

The agricultural contribution to the mitigation of climate change in terms of surface is appreciated by the UAA devoted to renewable energy.

Due to data availability, UAA devoted to renewable energy is limited to:

- ▶ areas of non-food set aside for energy generation (Reg (EC) 1251/1999)
- ▶ areas benefiting from the “Energy crop premium” (Reg (EC) 1782/2003)
- ▶ other areas devoted to energy crops without specific regime.

The latter can be estimated by balances, or derived from the production of bio fuel. This subdivision is important as new Member States opting for the Single Area Payment Scheme (currently all of them except MT & SI) were not obliged to set-aside. This transitional system ended in 2008. If available, UAA devoted to the production of short rotation coppice should also be taken into account.

26 – *Climate change/air quality: gas emissions from agriculture*

This indicator is measured as “Emissions of greenhouse gases and of ammonia from agriculture” – expressed in 1000 t of CO₂ equivalent for greenhouse gases or 1000 t of ammonia.

Greenhouse gases (GHG) as a whole include CO₂, CH₄, N₂O and fluorinated gases (HFCs, PFCs and SF₆).

According to the United Nations Framework Convention on Climate Change (UNFCCC) the following are sources of greenhouse gases from agriculture:

- i) enteric fermentation (CH₄);
- ii) manure management (CH₄, N₂O);
- iii) rice cultivation (CH₄);
- iv) agricultural soil management (CO₂, CH₄, N₂O);
- v) prescribed burning of savannahs (CH₄, N₂O); and
- vi) field burning of agricultural residues (CH₄, N₂O).

Emissions of GHG from land use change and forestry are excluded.

Carbon dioxide emissions do not include emissions from fossil fuel combustion sources that arise from agricultural-related processes such as transport, greenhouse heating and grain drying. Such sources are inventoried in IPCC under the Energy section, but the individual contribution of agriculture is not inventoried.

For GHG emissions, the primary source of data is the European Environment Agency. It compiles data received from the 25 Member States annual submission of data to the Secretariat of the United Nations Framework Convention on Climate Change (UNFCCC). Member States apply the 1996 IPCC guidelines to estimate the emissions and, they use the common reporting format (CRF) for submission of their inventories. Data collection via the EIONET (European Information and Observation Network) is being extended to include Candidate Countries which are becoming members of the European Environment Agency network.

For ammonia (NH₃) emissions, data are reported by Member States to the UNECE/EMEP Convention on Long-Range Transboundary Atmospheric Pollution (CLRTAP). Recommended methodologies for emission data collection are compiled in the Joint EMEP/CORINAIR Atmospheric Emission Inventory Guidebook (EMEP/EEA 2001).

Relevant Result Indicators

06 – Area under successful land management contributing to c) climate change

The indicator measures the total amount of hectares under successful land management. Successful land management is defined as the successful completion of land management actions contributing to : mitigating climate change

Information gathered from all these sources (bottom up) may be compiled to answer the following relevant evaluation questions:

5.4.2 Related Common Evaluation Questions and Horizontal Evaluation Questions according to Guidance Note B

Evaluation Questions

AXIS I, measure 121: Modernisation of agricultural holdings

To what extent have supported investments contributed to a better use of production factors on agricultural holdings? In particular, to what extent have supported investments facilitated the introduction of new technologies and innovation?

AXIS I, measure 123: Adding value to agricultural and forestry products

To what extent have supported investments contributed to introducing new technologies and innovation?

To what extent have supported investments contributed to enhancing market access and market share of agricultural and forest holdings, including sectors such as the sector of renewable energy?

AXIS I, measure 125: Improving and developing infrastructure related to the development and adaptation of agriculture and forestry

To what extent has the scheme contributed to restructuring and developing physical potential through the improvement of infrastructures?

AXIS II all measures where the Impact Indicator on Climate Change is identified as relevant (see Table 1 above)

AXIS II, measure 214: Agri-environment payments

To what extent have agri-environmental measures contributed to mitigating climate change?

AXIS II, measure 225: Forest environment payments

To what extent have forest-environment payments contributed to combating climate change?

AXIS III, measure 311: Diversification into non-agricultural activities

To what extent have supported investments promoted the diversification of farm households' activities towards non agricultural activities? Focus the analysis on the most important activities in this respect. To what extent have supported investments promoted additional employment opportunities for farm households outside the agricultural sector?

To what extent have supported investments contributed to improving the diversification and development of the rural economy?

To what extent have supported investments promoted the diversification of farm households' activities towards non agricultural activities? Focus the analysis on the most important activities in this respect. To what extent have supported investments promoted additional employment opportunities for farm households outside the agricultural sector?

AXIS III, measure 312: Support for business creation and development

To what extent has the support contributed to promote diversification and entrepreneurship? Focus the analysis on the most important activities.

To what extent has the support promoted additional employment opportunities in rural areas?

To what extent has the support contributed to improving the diversification and development of the rural economy?

Horizontal evaluation questions

To what extent has the programme contributed to promoting sustainable development in rural areas? In particular, to what extent has the programme contributed to the three priority areas for protecting and enhancing natural resources and landscapes in rural areas:

- ▶ climate change

To what extent has the programme integrated environmental objectives and contributed to the realisation of Community priorities in relation to

- ▶ the Kyoto protocol targets for climate change mitigation?

The following horizontal evaluation questions may show indirect (however relevant and significant) impact with relation to climate change:

To what extent has the programme contributed to further develop high quality and value added products?

To what extent has the programme contributed to promoting a strong and dynamic European agrifood sector?

To what extent has the programme contributed to promoting innovation in the European agrifood sector?

This list of evaluation questions shows quite clearly, that the range of RD measures potentially affecting Climate Change and thus producing a significant impact in this field is wide. It will call for a delicate balance to capture climate change impacts on the programme level on the one hand and the complexity of measuring and filtering deadweight and displacement effects on the other.

5.4.3 Key challenges with regard to measurement and interpretation

Measurement of this impact indicator as stipulated in the CMEF handbook is very well captured in the indicator fiche, following closely the induced output of ktOe installed by the support of RD programmes.

However the interpretation of this indicator as the RD programme contribution to combating climate change is limited. If relying only on this definition (of the impact on climate change), important aspects of climate change effects caused by RD interventions will not be captured – these are discussed below

5.4.3.1 Interpretation

What does the indicator explain and what not?

The current indicator relates to net greenhouse gas emissions reduction (i.e. carbon dioxide) attributable to the substitution of fossil fuels by non fossil alternatives such as

- ▶ Dedicated bioenergy crops: 'conventional' e.g. starch crops (e.g. cereals, sugar beets) or oil crops (e.g. rapeseed, sunflower); perennial grasses
- ▶ Short rotation forests on agricultural land
- ▶ Afforestation
- ▶ residues or biowaste (e.g. straw, greentops, manure)
- ▶ Wind and hydropower capacity

All these sources can displace emissions of fossil fuel energetic equivalent (KTOe)

While renewable energy production is a potentially useful indicator, it does not take into account other ways in which the RDP can impact on climate change. For example, it provides limited information on:

- ▶ Mitigation arising as a result of RDP-induced afforestation;
- ▶ Mitigation arising as a result of RDP-induced reductions in N-fertiliser application;
- ▶ Mitigation arising as a result of RDP-induced improvements in manure management;
- ▶ Mitigation arising as a result of RDP-induced changes in cultivation practice;
- ▶ Changes in the resilience of farms and their ability to *adapt* to climate change.

Many of these aspects can be remedied with reference to the broader baseline indicators, or the use of an additional indicator (CO₂e), which, as will be evident from existing country practice, is the broader interpretation of policy impact likely to be made by Member States.

Specific challenges regarding the interpretation of this indicator are outlined below.

Additionality/net effects: What proportion of the total new renewable energy capacity that has come on-stream on RDP recipient farms would not have happened in the absence of the RDP? To some extent, this can be determined with reference to the relative price of returns to growing biofuel and biomass crops relative to alternative land uses (i.e. conventional food crops). Inferior returns to energy crops suggests that support will be a key variable in decisions on switching and therefore additionality. In the case of investment in biogas or biomass combustion plants (heat and energy) there is a question of how assessment of impacts of RDPs can be isolated from the contribution of intervening domestic policies

When quantifying KTOe, we also need to identify the net effect after accounting for specific issues related to displacement (or leakage) of emissions, and spatial and temporal boundaries of the analysis.

Displacement of energy: What type of energy is the new source replacing? The extent to which new sources of renewable energy mitigate GHG emissions depend on the power generation that is displaced, e.g. an energy crop that displaces a kWh of electricity from a coal-fired power station will have a greater net reduction in emissions than an energy crop that displaces a kWh from a nuclear or hydro plant.

Displacement of production: Which farm activities have been displaced in order to enable the production of renewable energy? The measurement of the GHG mitigated by renewable energy production should take into account the effects of the land use change required to produce the renewable energy. For example, if an energy crop is grown, what land use has been lost, where will the lost (food) production take place, and what are the emissions associated with the displaced food production (i.e. including longer transportation and carbon intensity)? There is a risk that growing energy crops will simply displace production, and the associated emissions, to other places, which may or may not result in a net reduction in GHG emissions. To estimate the displacement effects, data are required on the land use and production changes brought about by the RDP-induced renewable energy.

This so-called “leakage” problem may be a judgement issue on where alternative foods come from. It also highlights the distinction between an evaluation processes that focuses on production rather than consumption emissions accounting (see below).

Boundary issues concern whether to limit consideration to the evaluation of impacts within the farm⁸⁴ boundary, (i.e. the KTOe represented by an area of energy crop), or whether to look at the ultimate net mitigation effects accounting the life-cycle costs of the energy source to its ultimate point of combustion. Thus, a field of coppice willow represents a notional amount of KTOe. But the ultimate net amount should account for added costs including transportation and processing prior to combustion (if the latter takes place off farm or the point of harvest)...

While their end use may be a beneficial displacement of fossil fuel alternative, the actual emissions benefit should realistically be net of their own production emissions. This is an important caveat to the current unit of account (ktoe).

There is also an issue of the temporal attribution of longer-term impacts to the policy period of the spending. This is because the mitigation potential of plant (e.g. biogas or wind) can take place for some

⁸⁴ note that farm can be inter changed with any other productive unit of land

decades beyond the programme period of the initial investment. An associated observation is that emissions are associated with an increasing marginal (i.e. unit) damage cost that can be undervalued by simply adding up tonnes of mitigation irrespective of when they occur. In other words, in terms of avoided damage costs, there is a social premium on earlier emissions reduction and this should be reflected in any evaluation.

One alternative is to convert these tonnes into their shadow price of carbon equivalent⁸⁵, estimates, which explicitly accounts for the increasing marginal damage of adding to an atmospheric stock. The avoided marginal damage cost attributable to an investment can then be discounted to a PV then annualised to fit inside the programme period. The fact that this can be done is another advantage of using the CO₂e metric rather than KTOe. However, this process means that already converting mitigation benefits into their monetary equivalent, which though possible and informative for efficiency purposes is not a stated requirement for this evaluation period.

5.4.4 Recommended methods of measurement

Describing micro and macro level and the aggregation from project to programme level

First it is important to distinguish between areas under fuel crops and investment in energy generation plant (e.g. biodigestors and biomass combustion). For fuel crop areas (and hence kTOe or CO₂e). The macro picture can be developed in two ways.

1. First from a bottom-up aggregation of measures that are applicable and interpreted as allowable under all Axes. This aggregation is in terms of KTOe and or CO₂e. mitigated from the application of identifiable measures associated with various policies across the axes. A bottom-up approach will require qualitative surveys of a cross section sample of recipients. Furthermore, this amount of CO₂e mitigation in aggregate must accord with data reported under IPCC inventory rules as attributable to agriculture land use land change. In other words, estimates of sink capacity (i.e. fuel crops) will be reported at a national scale as part of the inventory. Some part of this mitigation entry will be due to RDP support. This amount needs to be added to other Ktoe information related to energy generation from biogas. An aggregation of RDP-driven investment in these technologies can also be based on a bottom up survey and interviews (see the Austrian example below).
2. A more top down approach may be to use representative farm scale modelling to go from micro decisions to a macro or aggregate picture. It is possible to characterise a range of farm types using linear or dynamic programming methods, which can be used to evaluate land allocation decisions under alternative relative price changes. The use of representative modelling can then be a basis for evaluating aggregate conversion for given assumptions of support payments being received across a proportion of farms. The predicted areas can then be compared to a control counterfactual consisting of cropping decisions on a cross section of farms that are not given support. This modelling approach can then be compared with actual outcomes for a real sample. The latter information can be derived from the bottom-up approach.

⁸⁵ Example guidance on the shadow price of carbon is provided by UK Department of Energy and Climate Change (2009) Carbon Valuation in UK Policy Appraisal: A Revised Approach (July 2009) available www.decc.gov.uk Similar guidance is provided in other EU countries, e.g. France http://www.strategie.gouv.fr/article.php?id_article=830 "La valeur tutélaire du carbone".

Examples

Climate change mitigation is a relatively new metric in the context of the RDP, and there is difficulty in providing one route map for evaluation stages. However, since energy crops are simply an alternative land use, there is some commonality in terms of using farm accounts data for evaluating historical trends in cropping patterns and additionality in areas dedicated to cropping decisions under different relative price scenarios. Existing Member State experience is instructive about the different ways of interpreting how policy can have significant climate change impact through support to activity in all 4 axes. The key issue is to identify how relevant measures are activated within the range of support categories.

Current practices

Contrasting evaluation approaches are offered by Netherlands, Austria and Spain respectively. These examples were in agreement about the fact that climate change impacts can be delivered through measures in all three programme axes.

Example

The example from **Spain** demonstrates an approach that adheres rigorously to the CMEF interpretation of biofuel production and the KTOe indicator. The example (see next box below) combines a quantitative land use forecast (i.e. useable agricultural area) with qualitative interviews to evaluate a limited number of Axes 2 measures. The quantitative analysis considers the likely area to be recorded under miscellaneous biofuels. Interviews are used to understand the different crop choices and the extent to which these are likely to be used as substitutes for fossil fuels.

Netherlands and Austria both provide a wider interpretation of the evaluation approach by considering how a broader suite of practices that can be attached to programme measures in other axes (1,2 & 3) deliver climate change impacts.

The Austrian example demonstrates how to evaluate the displaced emissions from support to investment in agricultural biomass plants for electricity generation and heating. The impact of such energy measures are straightforward to evaluate in terms of KTOe by using specific assumptions about energy displaced. The savings are estimated from an assumed proportion of farm households assumed to substitute this generation technology for at least 50% of current generation. Additionality can be gauged by using a stratified sample of adopters to ascertain the true effect of the support.

The Austrian example also highlights how a range of policy-related agri environmental measures can be associated with practices that affect nitrogen use (thus its manufacture emissions and its loss from application) and methane from livestock. These impacts can only be reflected in terms of CO₂e, which by implication is what Austria intends to report alongside KTOe.

The Dutch experience is similar to the Austrian perspective, looking beyond Axis 2 and also considering how to measure impacts in terms of the more representative indicator CO₂e. The Dutch experience highlights the links to information contained in national greenhouse gas inventories collected for IPCC purposes. For example, IPCC Tier 1 data provides standard coefficients that can be used to quantify CO₂e impacts of quantitative changes in stocking rates or nitrogen application. The national inventory contains sufficient information to be useful for highlighting the impacts of a range of agricultural, land use and forestry practices. But the Dutch contribution also highlights the challenge in isolating RD programme support from other CC interventions.

Step by step process from baseline to impact

The following example from Spain provides a good overview how to tackle the assessment of the impact on climate change:

A. Assessment of the quantitative change in the production of renewable energy

In order to estimate the increase in production of potential renewable energy, the following sources have been taken into account:

1. Energy crops: cereals, beetroot, sunflower, sweet stalk corn.
2. Residues from agriculture: roots, leaves, straw, non-usable fruits, pruning residues.
3. Purpose-grown energy crops from forestry: poplars, eucalyptus, willow, acacia.
4. Residues from forestry: pruning residues.

1. Identification of related measures

According to the indicator description provided in the CMEF, the indicator addresses the following measures, 214, 216, 221, 222, 223, 224, 225, 226 and 227. To quantify the indicator, only the measures included in the evaluated RDP should be considered.

2. Assessment of the “positive impact area”

Based on the output indicators of supported land area or UAA (Utilised Agricultural Area) related to the selected measures mentioned above, the *positive impact area* (SPP) covered by each measure and/or commitment is calculated. This area is defined as the land, on which biomass for energy is expected to be produced.

It is estimated using coefficients to define the degree of influence of each measure and/or commitment on the indicator, within the context of the evaluated RDP (coefficient P_1 for measure 1, P_2 for measure 2, etc.). In order to avoid double-counting, overlapping between different measures and/or commitments has been taken into account, using weighing coefficients, as in the example that follows:

ST₁: Area supported by measure 1

ST₂: Area supported by measure 2

s₁₂: overlapping % of measures 1 and 2. $s_{12} = (ST_1 * ST_2) / (ST_1 + ST_2)^2$

Thus, the area covered only by measure 1 is estimated as $S_1 = (1 - S_{12}) * ST_1$, and the area covered by both measures 1 and 2 is calculated as $S_{12} = s_{12} * (ST_1 + ST_2)$. In the cases of overlapping measures, the coefficients which define the degree of influence of each measure are estimated as $P_{12} = 1 - [(1 - P_1) * (1 - P_2)]$.

Finally, the total area of positive impact for 17 indicator is the sum of $SPP = \sum (S_i * P_i)$.

3. Assessment of the potential biomass for energy production

Once the expected SPP is estimated, next step is to determine the yield, in terms of biomass for energy production. For this purpose, different energy crop utilization coefficients are used, depending on the type of crop or residue expected to be produced under each measure.

Depending on the source of the biomass, different calorific values are used to estimate the potential renewable energy kept within the kilograms of expected biomass.

Finally, the potential renewable energy is transformed into kilo tonnes of oil equivalent.

B. Assessment of the qualitative change in the production of renewable energy

The qualitative aspects considered to be related with the potential production of renewable energy are the following:

- Type of carbon sink, in terms of type of source (agriculture/forestry, purpose-grown crops/residues) and crop and/or tree species
- Type of actions, in terms of preventive actions (e.g. reducing forest fires), restoring actions (after fires and/or other disasters) and increasing actions
- Type of area/habitat supported, (mountain/handicap/normal areas, Natura2000/WFD/Special protected areas)
- Degree of use of the potential renewable energy produced within the RDP

In order to gather the information required to assess these aspects, case studies will be carried out among the recipients of the selected measures and/or commitments.

C. Drawbacks of the approach

1. Theoretical hypotheses

In the methodology described to estimate the expected value of the indicator at the end of the programming period, several hypotheses have been assumed, such as the following:

- Cultivated crops and/or trees in the supported land.
- Coefficients to define the degree of influence of each measure.

- Coefficients to estimate the overlapping degree between measures/commitments.

In the mid term evaluation a number of case studies will be selected in order to assess and quantify the impact indicator. Each year, during the programming period, the indicator will be monitored and the coefficients defined in the methodology will be checked and adjusted based on empirical data.

2. Addressed measures

According to the CMEF guidelines, a limited number of measures, within Axis 2, influence the indicator. Nevertheless, it is considered that impact indicators should be estimated at RDP level, as there are other measures that also have a direct influence on the indicator, at least the following:

121. Modernisation of agricultural Holdings

123. Adding value to agricultural and forestry products

125. Improving and developing infrastructure related to the development and adaptation of agriculture and forestry

126. Restoring agricultural production potential damaged by natural disasters and introducing appropriate prevention actions

211. Natural handicap payments to farmers in mountain areas

212. Payments to farmers in areas with handicaps, other than mountain areas

213. Natura 2000 payments and payments linked to Directive 2000/60/EC (WFD)

41. Implementing local development strategies with a view to achieving the objectives of one or more of the three other axes defined in sections 1, 2 and 3

421. Implementing cooperation projects (within past programming period several cooperation projects were directly related to renewal energy production and use, e.g. ELREN Project: European LEADER+ Renewable Energy Network, EURENERS Project: Energetic efficiency and promotion of renewal energies)

An additional indicator will be calculated, estimating the increase of potential renewal energy within the action of measures 211, 212 and 213. The approach applied will be as described above and the results will be summed up to the value of the common indicator.

Estimation of the effect of axis 1 measures is still under analysis.

Otherwise should we elaborate a bottom up or top down approach or both? These approaches then need to link to the text below

Additionality/programme net-effects: Assessing the proportion of the total new renewable energy capacity that has come on-stream on RDP recipient farms, which would not have happened in the absence of the RDP, can be addressed in several ways. The most basic approach is simply to ask payment recipients whether or not they would have produced the renewable energy if they had not received the RDP payment. This approach is not particularly reliable given the (perceived) risks to the respondent of losing future payments. Historical trends in the deployment of renewable energy may prove useful circumstantial evidence (provided data sets are available), however, they are unlikely to be able to provide compelling evidence of causal links between RDPs and renewable capacity due to the intervening variables such as energy prices. In order to control for intervening variables, a counterfactual scenario can be constructed using a control group of non-RDP recipients matched to the RDP recipients using farm datasets (FADN etc.). Alternatively, as mentioned, another approach would be to use farm scale modelling to assess the financial performance of different renewable energy investments on representative farm types. In the case of fuel crops this approach would allow the derivation of a support or factual “supply-response” schedule that could be compared to the counterfactual. For energy generation capacity farm modelling allows a comparison of, for example, the changes in gross margin on different farm types resulting from renewable energy investments with and without RDP support. Comparison of the “with and without” support cases would enable the theoretical uptake to be forecast and the difference in uptake with and without RDP support to be estimated. As always, the validity of the model results would depend on the assumptions and data input. In practice, all the methods mentioned above have some limitations, and a trade-off has to be made between the validity of the

estimate of additionality and the cost of estimating the additionality. Triangulation of results may be a useful way of improving the validity of the simpler methods.

Displacement of energy: previous sections have identified displacement issues, identifying and accounting for the energy generation displaced by the renewable energy. The calculation of the toe of additional renewable capacity should be adjusted to reflect the displaced energy by: (a) identifying the type of power generation displaced (b) multiplying the basic toe of renewable energy by a correction coefficient that reflects the relative carbon intensity of the renewable energy compared to the displaced source of energy. (a) should be identified via an examination of power generation fuel switching strategies, while the calculation of (b) should be relatively straightforward using national and EC energy statistics and basic thermodynamic data.

Displacement of production: Accounting for the farm activities displaced in order to enable the production of renewable energy. To estimate the displacement effects, data is required on the land use and production changes brought about by the RDP-induced renewable energy. In order to adjust: (a) estimate the area of land converted to renewable energy (using the data collected for Baseline Indicator 25 UAA devoted to renewable energy, and/or toe divided by the typical yield for the biomass); (b) estimate the yield foregone based on (a) and typical yields for the crops displaced; (c) identify where lost production will be sourced from; (d) qualitative assessment – is the change in production source likely to lead to a significant change in emissions? If so, this should be estimated and the toe adjusted to reflect the change.

Emissions arising from bioenergy production: The impact indicator should be adjusted to take account the emissions arising from the production of the bioenergy.

5.4.5 Data requirements and collection

Specific RDP evaluation experience on climate change is limited, but the measures above are dominated by those involving field cultivation (area/yields). This suggests that existing methods for determining area and yield additionally are applicable for energy crops. Methodologies include:

- ▶ Quantitative methods: Farm scale modelling scenarios, 'DiD' method and/or Propensity scoring of participation or land conversion under programme conditions relative to counterfactual
- ▶ Qualitative methods: farm interviews, reflective comparisons for farmers who are growing crops or who have installed capacity in biogas/wind/hydro

Data sources for all sources are relatively good

- ▶ Farm scale data on land under specific crops (FADN)
- ▶ Data can be partitioned in a variety of ways e.g. before/after programme; farm type and size
- ▶ Information on installed capacity (biogas and wind)
- ▶ Conversion factors to express biomass/gas/wind in terms of tonnes of oil equivalent
- ▶ Assumption: KTOe is measured at the farm gate (see below)

For wind/biogas generation, a distinction is made between energy use on farm and whether energy is exported to a national grid. For the latter the amount of renewable energy generated based on: (a)

Metering; (b) calculations based on installed capacity and average load factors; (c) where there aren't meters or a verified installed capacity

To determine this information, farmers and other land managers would be required to submit information of compliance with specific measures. This information could be subject to surveillance using systematic or randomised farm surveys and/or possible forms of triangulation using GIS or satellite data on land use. The same mitigation evaluation procedure could be incorporated into the evaluation procedure that is currently used for agri-environmental schemes.

Additional/Complementary indicators

In terms of mitigation, it is debatable whether or not the increase in the production of renewable energy is a sufficiently comprehensive measure of an RDP's contribution to combating climate change. In some cases the production of renewable energy represents only a small proportion of a farm's net GHG emissions. However, this is not necessarily a problem; indicators, by definition, only report a (usually) small measurable element of the parameter in question. The important point is whether or not changes in the indicator reflect changes in the bigger parameter. In the case of the climate change impact indicator, the amount of renewable energy is not related to the other abatement mechanisms, and therefore includes limited information about how the overall emissions from a particular farm might be changing in response to other measures within the RDP.

Example how to enrich climate change impact information

The following gives examples of the ways in which measures in the Scotland Rural Development Programme could abate GHG emissions.

Measures with potential to combat climate change in the Scotland Rural Development Programme 2007-13

Abatement mechanism	SRDP Measure with Abatement Potential
Reduction of the application of <i>nitrogenous fertilisers</i>	1/3. Nutrient Management Plan 1/5. Restructuring agricultural businesses? 2/33. Management of hedgerows? 2/34. Extended hedges? 2/35. Grass margins and beetlebanks 2/14. Management of species rich grassland 2/16. Creation and management of species rich grassland 2/18. Management of wetland 2/23. Buffer area for fens and lowland raised bogs 2/38. Management of ancient wood pasture 2/39. Scrub and tall herb communities
Manure management	1/3. Nutrient Management Plan 1/5. Restructuring agricultural businesses?
Reduce stock numbers	2/29. Moorland – stock disposal
Promotion of renewable energy	1/12. Processing and marketing of primary products 3/2. Support for rural energy – non-land based 1/5. Restructuring agricultural businesses? 2/33. Management of hedgerows 2/34. Extended hedges 3/1. Diversification outwith agriculture 3/3. Development/creation of micro-enterprises 2/46. Sustainable management of forests 2/47. Woodland improvement grant
Reduced cultivation	2/33. Management of hedgerows 2/34. Extended hedges 2/31. Off-wintering of sheep? 2/14. Management of species rich grassland 2/16. Creation and management of species rich grassland 2/38. Management of ancient wood pasture 2/39. Scrub and tall herb communities

Peat bog management	2/26. Wildlife management on upland and peat soils 2/28. Moorland grazings on upland and peatlands 2/20. Management/restoration of lowland raised bogs 2/30. Away-wintering of sheep 2/23. Buffer area for fens and lowland raised bogs
Afforestation/ Reforestation	2/45. Woodland creation 1/11. Improving the economic value of forests 1/14. Provision and upgrading of infrastructure 2/38. Management of ancient wood pasture 2/46. Sustainable management of forests 2/47. Woodland improvement grant

Numbers refer to the Axis and Measure, i.e. 1/3 = Axis 1, Measure 3. (Adapted from Moran et al. 2008)

The inclusion of an indicator such as *carbon dioxide equivalent* offers a more universal measure of the main gases causing climate change: methane, nitrous oxide and carbon dioxide. If this is an acceptable approach then there are potentially useful data (economising) links to inventory data used for IPCC reporting.

Carbon savings from forestry have been adopted as additional indicators some MSs (e.g. Scotland and England) and provide useful additional information. However, these indicators still fail to reflect the impact of measures that could reduce CO₂e in other ways, in particular via the potentially important mechanisms of reduced application of nitrogenous fertilisers and improved manure management. In order to test the validity of the impact indicators, the total potential mitigation that could be induced by the RDP measures could be calculated for representative farm types. This would show the proportion of the total climate change impact that is likely to be captured by the existing impact indicator, and different (sets of) additional indicators.

Finally, the CMEF climate change indicator focuses entirely on climate change mitigation (i.e. prevention of the release of emissions of greenhouse gas) with limited concern of adaptation to changing climatic conditions. This shortcoming is partially addressed in the Rural Development programme for England, through the use of the additional impact indicator "*Proportion of new woodland contributing to habitat networks and adaptation to climate change*".

5.4.6 Judgment issues

Cross referencing between environmental indicators

The delivery of climate change, water and HNMF outcomes all have implicit ancillary costs and benefits that need to be considered to derive a net picture of combined impact. Thus, targeting nitrogen in pursuit of water quality has inevitable impacts in terms of simultaneous reductions in atmospheric emissions and vice versa⁸⁶. Similarly, increased biomass and biofuel cropping will have implications for water demand, biodiversity outcomes and potentially food security. The nature of this environmental "cross-talk" is now being appreciated within policy formulation (e.g. within the WFD). Similarly, policies on ammonia reduction (principally for human health impacts) will also be relevant.

Current evaluation process does not suggest any requirement to adopt the most efficient mitigation practices first, particularly win-win measures⁸⁷. However, recent work on marginal abatement costs

⁸⁶ Targeting Nox emissions will likely deliver a diffuse pollution outcome.

⁸⁷ These are measures that simultaneously save farm costs and that bring emissions reductions – e.g. reducing excess Nitrogen application.

demonstrates that some measures are relatively efficient in terms of their cost per tonne of CO₂e mitigated⁸⁸. This UK working paper points out the undeveloped potential for efficiency in the adoption of practices through Axis measures

A final observation is that current evaluation focus is wholly focussed on production rather than consumption emissions. This is a general problem with current IPCC reporting protocol. The recent UK working paper refers to the need to consider consumer responses (in terms of labelling). Here we note that the general openness of agricultural trade means that a considerable amount of emissions displacement is associated with food production and consumption.

⁸⁸ <http://www.theccc.org.uk/pdfs/SAC-CCC;%20UK%20MACC%20for%20ALULUCF;%20Final%20Report%202008-11.pdf>

6. Bibliography

- ADE consulting, Pollen conseil and Edater : Ex-ante evaluation of the 2007-2013 hexagonal RDP
- Barca F. (2009): An Agenda for a Reformed Cohesion Policy. EC DG Regio, Bruxelles.
http://ec.europa.eu/regional_policy/policy/future/barca_en.htm
- Becu N., Perez P., Walker, A., Barreteau O. and Le Page C. (2004): Agent based simulation of a small catchment water management in northern Thailand: description of the CATCHSCAPE model. *Ecol. Model.* 170 (2–3), 319–331
- Behrendt, H., Huber, P., Ley, M., Opitz, D., Schmoll, O., Scholz, G., Uebe, R. (1999): Nährstoffbilanzierung der Flußgebiete Deutschlands. Berlin, UBA-Texte 79.
- Benoit M. (1992): Un indicateur des risques de pollution nommé BASCULE (Balance Azotée Spatialisée des systèmes de Culture de l'Exploitation). *Courier de la Cellule Environnement*, 18, 23–28.
- Benzler, Armin; Federal Agency for Nature Conservation (2009): *The implementation of the HNV farmland indicator in Germany*. Rural Evaluation News no. 2.
- Biewinga E.E., van der Bijl G. (1996): Sustainability of Energy Crops in Europe: A Methodology Developed and Applied. Centre for Agriculture and Environment, Utrecht, The Netherlands, p. 209.
- Blien, U. Suedekum, J., O'Wolf, K. (2005): Local employment growth in West Germany: A dynamic panel data approach. IZA Discussion Paper No 1723. www.iza.org
- Bradley, J., Gács, J., Kangur, A., Lubenets, N. (2003): 'Macro impact evaluation of National Development Plans: A tale of Irish, Estonian and Hungarian collaborations' Paper presented at the Fifth European Conference on Evaluation of the Structural Funds, Challenges for evaluation in an Enlarged Europe, Budapest.
- Caliendo, M., Kopeinig, S. (2008) 'Some practical guidance for the implementation of propensity score matching', *Journal of Economic Surveys* 22(1): 31-72.
- Chapanis, A. (1959): Research techniques in human engineering, John Hopkins Press, Baltimore: Maryland.
- Checkland, P.B. (2001): Soft Systems Methodology, in J. Rosenhead and J. Mingers (eds), *Rational Analysis for a Problematic World Revisited*. Chichester (Wiley)
- CMEF Guidance document (including all Guidance Notes A-O)
http://ec.europa.eu/agriculture/rurdev/eval/guidance/document_en.pdf
- Cristofini B. (1985): La petite région vue au travers du tissu de ses exploitations. Un outil pour l'aménagement et le développement rural. *Etud. Rech. Syst. Agraires Dév.* 6, 44.
- de Koning G.H.J., van de Kop, P.J., Fresco, L.O. (1997): Estimates of sub-national nutrient balances as sustainability indicators for agro-ecosystems in Ecuador. *Agric. Ecosyst. Environ.* 65 (2): 127–139.
- DG Agriculture (2004): Impact assessment of rural development programmes in view of post 2006 rural development policy. A Report submitted by EPEC, Brussels.
- DG Agriculture (2007): Handbook on Common Monitoring and Evaluation Framework – Guidance Note A

ECNC (2000). In: Wascher, D.W. (Ed.), *Agri-environmental Indicators for Sustainable Agriculture in Europe*. European Center for Nature Conservation, Tilburg, The Netherlands, p. 240.

ECORYS (2005): *Impact analysis: study on baseline and impact indicators for rural development programming 2007-2013*. Synthesis Report AGRI/2004/G2/02.

EENRD, EC (2008): *Guidance Document – The Application of the High Nature Value Impact Indicator*. Available on the web at http://ec.europa.eu/agriculture/rurdev/eval/hnv/guidance_en.pdf and http://ec.europa.eu/agriculture/rurdev/eval/hnv/guidance_annexes_en.pdf

European Bird Census Council (2004): *European wild bird indicators, 2008 update*
<http://www.ebcc.info/index.php?ID=368>

Eurostat (2009): *Labour force survey in the EU, candidate and EFTA countries*. This publication describes the main characteristics of the national surveys. Eurostat also produces working papers on the topic: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-08-007/EN/KS-RA-08-007-EN.PDF

Eurostat (2009): *Labour force survey in the EU, candidate and EFTA countries*. This publication describes the main characteristics of the national surveys. Eurostat also produces working papers on the topic: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-08-007/EN/KS-RA-08-007-EN.PDF

EVALSED (2008): *The Guide*. URL:
http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/guide/methods_techniques/indicators/assessing_en.html

Fitschen-Lischewski A. (2009): *Measurement of the Indicator Labour Productivity*. Paper presented at the Workshop on Assessing the Socio-Economic Impacts of Rural Development Programmes (Roma, 27/10/2009).

Forrester J.W. (1971): *Principles of systems*; Wright-Allen Press; Cambridge, Mass. 2. preliminary ed., 5. print.

Fransella, F., Bell, R., Bannister, D. (2004): *A Manual for Repertory Grid Technique*. 2. ed. Chichester (Wiley)

Gadamer, H-G. (1979): *Truth and Method*, London: Sheed and Ward.

Galloway, H.: *EMSI's Input-Output Model Multipliers: A Brief Overview and Comparison with Other Major Models*, www.economicmodeling.com

Geier U., Kopke, U. (1998): *Comparison of conventional and organic farming by process-life cycle assessment. A case study of agriculture in Hamburg*. In: *Proceedings of the International Conference on Life Cycle Assessment in Agriculture, Agroindustry and Forestry*, Brussels, Belgium, pp. 31–38.

Giupponi C., Eiselt B., Ghetti P.F. (1999): *A multi criteria approach for mapping risks of agricultural pollution for water resources: the Venice Lagoon watershed case study*. *J. Environ. Manage.* 56 (4): 259–269.

GTZ (2003): *Establishing Plausibility in Impact Assessment*, Workgroup on Assessing the Impact of Agricultural Research in Development, GTZ, Eschborn.

Hellenic Ornithological Society (2008): *Identification of high nature value agricultural and forestry land – executive summary (final report)*. Ministry of Rural Development and Food Operational Programme: Agricultural Development and Reform of the Countryside.

Helpdesk of the European Evaluation Network for Rural Development (2009): *Note regarding the organisation of activities I.1.1 and I.1.2 of the Annual Work Programme 2009 as a joint thematic work*

group: "Approaches for assessing the impacts of the rural development programmes in the context of multiple intervening factors." Bruxelles.

Hengsdijk H., van Ittersum M.K. (2003): Formalizing agro-ecological engineering for future-oriented land use studies. *Eur. J. Agron.* 19 (4), 549–562.

Henrichsmeyer, W., Cypris, C., Löhe, W., Meudt, M., Sander, R., von Sothen, F., Isermeyer, F., Schefski, A., Schleef, K.-H., Neander, E., Fasterding, F., Helmcke, B., Neumann, M., Nieberg, H., Manegold, D., Meier, T. (1996): Entwicklung eines gesamtdeutschen Agrarsektormodells RAUMIS96. Endbericht zum Kooperationsprojekt. Forschungsbericht für das BML (94 HS 021), vervielfältigtes Manuskript Bonn/Braunschweig.

Hoffmann-Kroll, R., Benzler, A., Schäfer, D. and Seibel, S. (2000): Setting-up national biodiversity monitoring for nature conservation in Germany – the Ecological Area Sampling (EAS). In Bischoff, C. & Dröschmeister, R. (eds): European Monitoring for Nature Conservation. Schriftenr. Landschaftspflege Naturschutz 62, 79-94.

Hummelbrunner R. (2005): Process Monitoring of Impacts. Towards a New Approach to Monitor the Implementation of Structural Funds Programmes. ÖAR Regionalberatung, Wien

IEEP (2007): HNV Indicators for Evaluation, Final report for DG Agriculture, (Contract Notice 2006-G4-04), IEEP, London.

<http://ec.europa.eu/agriculture/analysis/external/evaluation/guidance.pdf>

Janetschek H. (2009): Methods suitable to evaluate investment support. Paper presented at the Workshop on Assessing the Socio-Economic Impacts of Rural Development Programmes (Roma, 27/10/2009).

Jørgensen B. (1993): The theory of linear models; Chapman & Hall; New York, NY

Kreins P., Gömann H., Heidecke C., Hirt U., Richmann A., Seidel K., Tetzlaff B., Wendland F. (2009): Costs of achieving objectives of the WFD by reducing diffuse nitrogen leaching in agriculture in the Weser river basin. Paper to be presented at the 8th International Conference of the European Society for Ecological Economics, Ljubljana, Slovenia 29 June to 2 July 2009 (attached as pdf-file)

Kreins P., Gomann H., Heidecke C., Hirt U., Richmann A., Seidel K., Tetzlaff B., Wendland F. (2009): Costs of achieving objectives of the Water Framework Directive by reducing diffuse nitrogen leaching in agriculture in the Weser River Basin. *ESEE*, in press.

Laschewski, L.; Schmidt, C. (2008): The utilization of evaluation results in the design of EAFRD – a study of practices of European agricultural policy making. In: Laschewski, L. (Hrsg.): Ländliche Gesellschaft und ländliche Politik als Prozess. Formation von Sozialkapital, Demografischer Wandel, Politische Entscheidungen, Forschungsberichte 2/2008, Universität Rostock, pp. 52-80.

Le Roux, X., Barbault, R., Baudry, J., Burel, F., Doussan, I., Garnier, E., Herzog, F., Lavelle, S., Lifran, R., Roger-Estrade, J., Sarthou, J.P., Trommetter, M. (éditeurs) (2008). *Agriculture et biodiversité. Valoriser les synergies*. Expertise scientifique collective, synthèse du rapport, INRA (France).

Leeuw F., Vaessen J. (2009): Impact Evaluations and Development. NONIE Guidance on Impact Evaluation. NONIE, Washington. <http://www.worldbank.org/ieg/nonie/guidance.html>

Leip A., Marchi G., Koeble R., Kempen M., Britz W., Li C. (2007): Linking an economic model for European agriculture with a mechanistic model to estimate nitrogen losses from cropland soil in Europe. *Biogeosciences Discussion*, 4: 2215–2278.

- Leuven, E., Sianesi, B. (2003): PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. RePEc, Research Papers in Economics: <http://ideas.repec.org/c/boc/bocode/s432001.html>. Accessed 4.12.2006.
- Lukesch R., Payer H., Rabenau J. (2005): Wissen von innen – Fokusgruppen in der Begleitforschung zu Regionen Aktiv. In: Böcher M., Krott M., Tränkner S. (2005): Regional Governance und integrierte ländliche Entwicklung. Ergebnisse der Begleitforschung zum Demonstrationsvorhaben „Regionen Aktiv“. VS Verlag für Sozialwissenschaften, Wiesbaden
- Moran, D., Anita Wreford, Michael MacLeod, Eileen Wall, Julian Bell, Caroline McBain, Gavin Dick (2008): ASSESSING THE POTENTIAL FOR SRDP MEASURES TO MITIGATE AGAINST CLIMATE CHANGE AA2 I I study for RERAD Edinburgh: Scottish Government
- OECD/Eurostat (2003): OECD/Eurostat Gross Nitrogen Balances Handbook, <http://webdomino1.oecd.org/comnet/agr/aeiquest.nsf> (password required).
- OIR (2009): TERESA project (project conducted in the 6th RFP – <http://www.teresa-eu.info/>).
- ÖIR-Managementdienste GmbH (2003): Ex-post evaluation of the Community Initiative LEADER II, http://ec.europa.eu/agriculture/eval/reports/leader2/index_en.htm
- Panten K, Rogasik J, Godlinski F, Funder U, Greef J-M, Schnug E (2009): Cross soil surface nutrient balances: The OECD approach implemented under German conditions. *Landbauforsch* 59(1):19-28
- Papy F. (2001): For a theory of field management: an agronomy of territories. *C.R. Acad. Agric. Fr.* 87 (4), 139–149.
- Paracchini, M. L., Petersen, J-E., Hoogeveen, Y., Bamps, C., Burfield, I. and van Swaay, C. (2008): High Nature Value Farmland in Europe. An Estimate of the Distribution Patterns on the Basis of Land Cover and Biodiversity Data. European Commission Joint Research Centre, Institute for Environment and Sustainability. Office for Official Publications of the European Communities, Luxembourg, accessed through http://agrienv.jrc.it/publications/pdfs/HNV_Final_Report.pdf.
- Parson, L. S. (2001): Reducing bias in a propensity score matched-pair sample using greedy matching techniques. Proceedings of the 26th Meeting of the SAS International User Group (SUGI 26): <http://www2.sas.com/proceedings/sugi26/p214-26.pdf>. Accessed: 5.10.2006.
- Payraudeau S., van der Werf H.M.G. (2005): Environmental impact assessment for a farming region: a review of methods. *Agriculture, Ecosystems and Environment* 107: 1–19.
- Petit O., Corcos B., O'Connor, M. (2001): Appropriation Sociale du problème de gestion durable de la nappe de Beauce. Programme Inter-institutionnel de Recherches et d'Études en Economie de l'Environnement (PIREE) du Ministère de l'Aménagement du Territoire et de l'Environnement. Université de Versailles-Saint Quentin en Yvelines, Guyancourt, France, p. 90.
- Petrick, M., Zier, P. (2009): Employment impacts of the Common Agricultural Policy in Eastern Germany – A regional panel data approach. *Research in Agricultural and Applied Economics*: <http://agecon.lib.umn.edu/>.
- Poux X., Romain B. (2009): L'agriculture à Haute Valeur Naturelle: mieux la (re)connaître pour mieux l'accompagner. European Forum on Nature Conservation and Pastoralism. www.efncp.org
- Pufahl, A. (2009): Studie I: Einkommens- und Beschäftigungswirkungen von Agrarumweltmaßnahmen, der Ausgleichszulage für benachteiligte Gebiete und der Ausgleichszulage für Gebiete mit

umweltspezifischen Einschränkungen. In: LR, Institut für Ländliche Räume des Johann Heinrich von Thünen-Instituts (Hrsg.): Ex-post-Bewertung des Hessischen Entwicklungsplans für den Ländlichen Raum. Materialband zu Kapitel 10 Kapitelübergreifende Fragestellungen. Braunschweig.
<http://www.vti.bund.de/de/institute/lr/publikationen/default.htm>. Accessed: 7.10.2009.

Pufahl, A., Weiss, C. R. (2009): Evaluating the effects of farm programs: Results from propensity score matching. *European Review of Agricultural Economics* 36 (1), 79-101.

Rasul G., Thapa G.B. (2004): Sustainability of ecological and conventional agricultural systems in Bangladesh: an assessment based on environmental, economic and social perspectives. *Agric. Syst.* 79 (3), 327–351.

Republic of Bulgaria (2007): *Rural Development Programme (2007-2013)*, Annex 3. December 2007 version.

Rodrigues G.S., Campanhola C. and Kitamura P.C. (2003): An environmental impact assessment system for agricultural R&D. *Environ. Impact Assess. Rev.* 23 (2), 219–244.

Schroder J.J., Aarts H.F.M., ten Berge H.F.M., Van Keulen H., Neeteson J.J. (2003): An evaluation of the whole-farm nitrogen balances and related indices for efficient nitrogen use. *European Journal of Agronomy*, 20: 33-44.

Schulze E.D., Luyssaert P., Ciais P., Freibauer A., Janssens I.A. et al. (2009): Importance of methane and nitrous oxide for Europe's terrestrial greenhouse-gas balance. *Nature Geosciences*, 22 November 2009, DOI: 10.1038/NGE686.

Treyz, F., Treyz, G. I. (2004): 'The Evaluation of Programs aimed at Local and Regional Development: Methodology and Twenty Years of Experience using REMI Policy Insight', in OECD (ed) *Evaluating Local Economic and Employment Development – How to assess what works among programmes and policies*, Paris: Organisation for Economic Co-operation and Development – LEED Programme.

UK Department of Energy and Climate Change (2009): Carbon Valuation in UK Policy Appraisal: A Revised Approach (July 2009) available www.decc.gov.uk. Similar guidance is provided in other EU countries, e.g. France http://www.strategie.gouv.fr/article.php3?id_article=830 "La valeur tutélaire du carbone".

UN (1993): System of National Accounts 1993. Online Resource at:
<http://unstats.un.org/unsd/sna1993/introduction.asp>

Van Dingenen, Katrien (IDEA-Consult, 2009): Current Practice Example presented during European Evaluation Network workshop on assessment of socio-economic indicators, Rome, 28 October 2009

van Eerd M.M., Fong P.K.N. (1998): The monitoring of nitrogen surpluses from agriculture. *Environmental Pollution*, 102: 227-233.

Velthof G.L., Oudendag D., Witzke H.P., Asman W.A.H., Klimont Z. & Oenema O. (2009): Integrated assessment of nitrogen losses from agriculture in EU-27 using MITERRA-EUROPE. *Journal of Environmental Quality*, 38: 402-417.

Velthof G.L., Oudendag D., Witzke H.P., W.A.H. Asman W.A.H., Z. Klimont Z., Oenema O. (2009): Integrated Assessment of Nitrogen Losses from Agriculture in EU-27 using MITERRA-EUROPE. *J. Environ. Qual.* 38:402–417.

Verburg P.H., van de Steeg J., Veldkamp A., Willemen L. (2009): From land cover change to land function dynamics: A major challenge to improve land characterization. *Journal of Environmental Management*, 90: 1327-1335.

Watzlawick P. (1977): *How real is real? Confusion, Disinformation, Communication*. Vintage Publications, Millers Falls, MA.

Wendland, F., Kunkel, R., Grimvall, A., Kronvang, B., Müller-Wohlfeil, D. I. (2002): The SOIL-N/WEKU model system – a GIS-supported tool for the assessment and management of diffuse nitrogen leaching at the scale of river basins. *Water Science and Technology*, 45 (9): 285-292.

Wendland, F., Kunkel, R., Voigt, H.-J. (2004): Assessment of groundwater residence times in the pore aquifers of the River Elbe Basin. *Environmental Geology*, 46: 1-9.

Wooldridge, J. M. (2002): *Econometric estimation of cross-sectional and panel data*. Cambridge, MIT Press.

Worldbank (2008): *Ukraine: Agricultural Competitiveness*. Europe and Central Asia Region Sustainable Development Unit.

Yin, R. K. (2008): *Case Study Research: Design and Methods*, 4th Edition, Thousand Oaks: Sage Publications. Everyone conducting case studies should have read this.

Zander P., Kachele H. (1999): Modelling multiple objectives of land use for sustainable development. *Agric. Syst.* 59 (3): 311–325.

Internet Sources for assessing and evaluating impacts

Evaluation Designs.

<http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPM/0,,contentMDK:20188242~menuPK:415130~pagePK:148956~piPK:216618~theSitePK:384329,00.html#sb>

Impact Evaluation Methods.

<http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/AFRICAEXT/EXTIMPEVA/0,,contentMDK:21017034~menuPK:2830194~pagePK:64168445~piPK:64168309~theSitePK:2620018,00.html>

Estimation Methods.

<http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPM/0,,contentMDK:20188244~menuPK:384339~pagePK:148956~piPK:216618~theSitePK:384329~isCURL:Y,00.html>

HNV Guidance Document – see link

http://ec.europa.eu/agriculture/rurdev/eval/hnv/guidance_en.pdf

Impact evaluation: methods and implementation issues.

[www.ifc.org/ifcext/sme.nsf/AttachmentsByTitle/MEImpactEvaluation.pdf/\\$FILE/MEImpactEvaluation.pdf](http://www.ifc.org/ifcext/sme.nsf/AttachmentsByTitle/MEImpactEvaluation.pdf/$FILE/MEImpactEvaluation.pdf)

Impact Evaluation. Methodological and Operational Issues.

<http://www.adb.org/Documents/Handbooks/Impact-Analysis/default.asp>

Experimental and quasi-experimental techniques.

http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/downloads/sb2_experimental_quasi_methods.doc

Program Evaluation Methods: Measurement and Attribution of Program Results. www.tbs-sct.gc.ca/eval/pubs/meth/pem-mep_e.pdf

Conducting Quality Impact Evaluations Under Budget, Time and Data Constraints.

[http://lnweb90.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/757A5CC0BAE22558852571770059D89C/\\$file/conduct_qual_impact.pdf](http://lnweb90.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/757A5CC0BAE22558852571770059D89C/$file/conduct_qual_impact.pdf)

Propensity score-matching methods for non-experimental casual studies.

www.nber.org/~rdeheja/papers/matching.pdf

Quasi-experimental and experimental approaches to environmental economics.

http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6WJ6-4TN0KS5-1&_user=10&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=077ad6ef6d630897b54a9256c9f56d4f

Nonie Sub-group 1: Experimental and quasi-experimental approaches to impact evaluation.

www.worldbank.org/ieg/nonie/docs/subgroup1.doc

Difference-in-Difference Estimation. www.nber.org/WNE/lect_10_diffindiffs.pdf

Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results.

<http://www.gsdr.org/docs/open/EIRS4.pdf>

Qualitative Research Methods. <http://projects.exeter.ac.uk/prdsu/helpsheets/Helpsheet09-May03-Unlocked.pdf>

Qualitative methods. <http://www.socialresearchmethods.net/kb/qualmeth.php>

7. Abbreviations

AE	Agri-Environment
AES	Agri-Environmental Schemes
AFP	Agrarinvestitionsförderungsprogramm
ATE	Average Treatment Effects
ATNT	population average treatment effect on non-treated
ATT	Average Treatment on Treated
ATU	Average Treatment on Untreated
CAP	Common Agricultural Policy
CEQ	Common Evaluation Question
CGE	Computable General Equilibrium
CLRTAP	Convention on Long-Range Transboundary Atmospheric Pollution
CMEF	Common Monitoring and Evaluation Framework
CORINE	COoRdinate INformation of the Environment (European Environment Agency Land Cover database)
DD	Double Difference
DID	Difference In Difference
EAFRD	European Agricultural Fund for Rural Development
EBCC	European Bird Census Council
EFTA	European Free Trade Association
EIONET	European Information and Observation Network
EUR	Euro
FADN	Farm Accountancy Data Network
FBI	Famland Bird Indicator
FSS	Farm Structure Survey
FTE	Full Time Equivalent
GE	General Equilibrium
GHG	GreenHouse Gases
GNB	Gross Nutrient Balance
GPS	Generalized Propensity Score matching
GVA	Gross Value Added
HNV	High Nature Value
KToe	1,000 tons of oil equivalent
ktons	1,000 tons
LEADER	Actions for innovative rural development (Community Initiative)
LFA	Less Favoured Area
LPIS	Land Parcel Identification System
LSDV	Least Square Dummy Variable
MCPFE	Ministerial Conference on the Protection of Forests in Europe
MEUR	million euros
MS	Member State
NGO	Non-Governmental Organisation
NUTS	« Nomenclature des unites territoriales » (= Nomenclature of Territorial Units for Statistics; Geocode standard for referencing the administrative division of countries for statistical purposes.)

OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Square
PATE	Population Average Treatment Effect
PATT	Population Average Treatment effect on Treated
PCA	Principal Component Analysis
PECBM	Pan-European Common Bird Monitoring scheme
PPP	Purchasing Power Parity
PPS	Purchasing Power Standard
PSM	Propensity Score Matching
R&D	Research and Development
RD	Rural Development
RDD	Regression Discontinuity Design
RDI	Rural Development Index
RDP	Rural Development Programme
SAPARD	Special Accession Programme for Agriculture and Rural Development
SATE	sample average treatment effect
SATT	sample average treatment on treated
SME	Small and Medium-sized enterprises
STNT	sample effect on non-treated
SWOT	Strengths, Weaknesses, Opportunities, Threats
UAA	Utilised Agricultural Area
UNFCCC	United Nations Framework Convention on Climate Change
UR	Unit of Reference