



**HAL**  
open science

## Comparison of methods to assess the sustainability of agricultural systems. A review

Christian C. Bockstaller, Laurence Guichard, Olivier Keichinger, Philippe Girardin, Marie-Béatrice Galan, Gérard Gaillard

► **To cite this version:**

Christian C. Bockstaller, Laurence Guichard, Olivier Keichinger, Philippe Girardin, Marie-Béatrice Galan, et al.. Comparison of methods to assess the sustainability of agricultural systems. A review. *Agronomy for Sustainable Development*, 2009, 29 (1), pp.223-235. 10.1051/agro:2008058. hal-01173169

**HAL Id: hal-01173169**

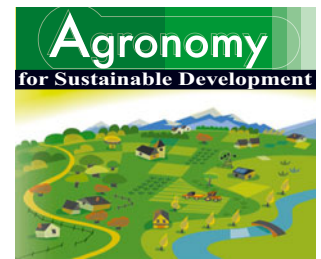
**<https://hal.science/hal-01173169>**

Submitted on 31 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright



## Review article

# Comparison of methods to assess the sustainability of agricultural systems. A review

Christian BOCKSTALLER<sup>1\*</sup>, Laurence GUICHARD<sup>2</sup>, Olivier KEICHERGER<sup>1</sup>, Philippe GIRARDIN<sup>1</sup>,  
Marie-Béatrice GALAN<sup>3</sup>, Gérard GAILLARD<sup>4</sup>

<sup>1</sup> INRA, UMR 1121 Nancy-Universités – INRA Agronomie et Environnement Nancy-Colmar, BP 20507, 68021 Colmar Cedex, France

<sup>2</sup> INRA, UMR 211 Agronomie INRA-AgroParisTech, Bâtiment EGER, BP 01, 78850 Thiverval-Grignon, France

<sup>3</sup> Agro-Transfert Ressources et Territoires Picardie, 80200 Estrées-Mons, France

<sup>4</sup> Agroscope Reckenholz-Tänikon Research Station ART, Reckenholzstrasse 191, 8046 Zurich, Switzerland

(Accepted 3 October 2008)

**Abstract** – Since the 1990s, numerous agri-environmental indicators and indicator-based methods have been developed to assess the adverse effects of cropping and farming systems such as water pollution by nitrates and pesticides, and gaseous emissions due to nitrogen inputs. This wealth of environmental indicators and assessment methods based on indicators raises issues on the quality of the methods and of the indicators, and on the relevancy of results. Evaluation and comparative studies are therefore needed to answer such issues. Here, we present four recent comparative studies selected for their illustrative value, first, to analyse the methodologies used for comparison of methods, and second, to highlight the main results of the four comparisons. The first study involves 23 indicators to address nitrate leaching. The second study involves 43 indicators to address pesticide risk. The third and fourth studies compare environmental assessment methods based on 4–5 indicators used in French and Upper Rhine plains (France, Germany and Switzerland). Both studies also compare the outputs of the methods and highlight the low degree of convergence between them. The approach proposed in the last study is the most elaborate among the four case studies. It could be used to develop a generic evaluation and comparison methodology. The review of those four case studies shows the need to formalise the methodology underlying any comparison work of indicators or evaluation methods.

**environmental assessment / indicators / nitrogen / pesticide / nitrate**

## 1. INTRODUCTION

During the '90s, there was a growing concern for environmental issues in agriculture, e.g. water pollution by nitrates, pesticides, erosion, or more recently, greenhouse gas emissions and biodiversity losses (Kirchmann and Thorvaldsson, 2000). This led to the demand for operational assessment tools considered as a prerequisite to the development of new farming or cropping systems (Bockstaller et al., 1997). This was favoured by the popularisation of the concept of environmental management approaches like the ISO-14000 which rest on the four steps of the “quality spiral of continuous process improvement”: to plan, to do, to check, to act (Meynard et al., 2002). The step “check” requires an assessment method of environmental impacts. The use of indicators has appeared as an alternative to direct impact measurement (Mitchell et al.,

1995; Bockstaller et al., 2008), and is linked to methodological difficulties (impossibility of measurement, complexity of the system) or practical reasons (time, costs) carrying out direct measurements. Another reason is the use of such tools for prospective goals (development of new agricultural strategies, prevention of environmental damage) in an ex ante assessment for which it is per definition not possible to perform measurements.

An “indicator explosion” (Riley, 2001a) has occurred for the last two decades with the development of numerous indicator-based methods which are aimed at assessing environmental impacts of agricultural activities, or the whole sustainability of agricultural systems (Rosnoblet et al., 2006). Among the works on indicators, one can distinguish those on a specific theme, on one hand, like the environmental risk due to pesticide application (Maud et al., 2001; Reus et al., 2002; Devillers et al., 2005) and multi-criteria approaches based on

\* Corresponding author: [bockstal@colmar.inra.fr](mailto:bockstal@colmar.inra.fr)

a set of indicators addressing different environmental issues, on the other. Examples at regional (Payraudeau and van der Werf, 2005), farm (Eckert et al., 2000; van der Werf and Petit, 2002; Hülsbergen, 2003; Meyer-Aurich, 2005), and cropping system levels (Bockstaller et al., 1997; López-Ridaura et al., 2005; Nemecek et al., 2005) can be given for both.

This multiplicity and variety of indicators and methods raise questions. Riley (2001b) pointed out that it is a source of confusion which is increased by the fact that many methods are not evaluated for their scientific relevance and feasibility. The potential user, either a researcher working on innovative cropping systems, or an adviser working with farmers or a stakeholder involved in an environmental debate, will have questions about the selection of a given method adapted to his needs and how to make this selection. A second group of questions deals with the stability of the outputs of the different methods: do they provide the same conclusions? Answers to such questions require an evaluation and comparison study which provides information, not only about the strengths and drawbacks of each method, its field of use and validity, but also about the comparison of the conclusions derived from the outputs of the methods. Some authors (Meynard et al., 2002; Bockstaller et al., 2008) have already pointed out the requirement of a comparative analysis and validation of the various indicators available. To answer this concern of potential users, different kinds of comparative works have been undertaken. Comparison works of assessment methods based on a set of indicators, such as those at farm level (van der Werf and Petit, 2002; Halberg et al., 2005) or regional level (Payraudeau and van der Werf, 2005) are based on a descriptive approach. In other comparative studies on impact assessment (Thompson, 1990; Hertwich et al., 1997) or more specific to the agricultural sector (Gebauer and Bäuerle, 2000; Thomassen and de Boer, 2005), authors use a set of qualitative or semi-quantitative evaluation criteria to compare the methods. No information is given on the comparison of the outputs or conclusions of the methods by all those authors, except by Thomassen and de Boer, 2005. They also study correlation between results of comparable indicators belonging to the input-output accounting approach and Life Cycle Analysis for a dataset obtained on 8 dairy farms.

This short review of the literature points out the diversity of approaches and a lack of formalised comparison methodology. The first goal of this article is to analyse the methodologies used in four comparative studies (CORPEN, 2006; Devillers et al., 2005; Galan et al., 2007; Bockstaller et al., 2006), selected to derive some methodological principles for potential users who need to undertake such a comparison. Second, the main results of the four comparisons will be highlighted to guide potential users of indicators or an evaluation method in their choice. Attention is paid to agronomists working on the design of innovative cropping systems and to environmental impact due to pesticides and nitrogen issues, for which many indicators are available. The four case studies structuring the article were selected for their diversity and illustrative value. The type of indicators and methods covered by the case studies and their target users, agronomists assessing and designing cropping systems, was another reason for their selection.

## 2. PRESENTATION OF THE FOUR CASE STUDIES: CONTEXT AND METHOD OF COMPARISON

### 2.1. Comparison of indicators assessing nitrogen losses

#### 2.1.1. Context of the work

The work was initiated by the CORPEN organisation, which depends on the French Ministry for Ecology and Sustainable Development and has the mission to bring together experts and stakeholders involved in the issue of water quality and agriculture in order to deliver recommendations (CORPEN, 2006). The objective was to help users to choose and to implement indicators depending on the question and the scale of study. It was carried out by a group of experts on nitrogen fertilisation and losses, from research and technical institutes. The expert group listed 23 indicators currently used by agricultural advisors, farmers or even public policy agents to assess nitrogen losses in France, and especially nitrate leaching at farm and regional levels. For the sake of concision, we will restrict the presentation of this work to the field and farm scale since our article addresses the evaluation of cropping and farming systems.

#### 2.1.2. Method of comparison

A descriptive sheet was filled in for each indicator with a list of descriptors: reference values, calculation method, time and spatial scale, periodicity of calculation, time for implementation, recommendations for interpretation and similar indicators, etc. In the report of the project, a synthetic table was added to present the assessment of two evaluation criteria for 15 indicators: (1) the feasibility, i.e. easiness of implementation due to accessibility of data and cost of implementation expressed on a qualitative scale between 1 (low) and 4 (high), and (2) the relevance assessed by experts on a four-class scale, from 1 (indicator not to be implemented alone) to 4 (indicator “advised”). Indicator sheets as well as the two evaluation criteria were filled in by members of the group of experts and validated by the group of experts. A selection of descriptors and the assessment of the two criteria are presented in Table III in Section 3.1 of the “results” chapter.

### 2.2. Comparison of 43 pesticide risk indicators

#### 2.2.1. Context of the work

This work followed the studies of Maud et al. (2001) and Reus et al. (2002) who compared, respectively, 6 and 8 pesticide risk indicators. The study was ordered by the French Ministry for Ecology and Sustainable Development and was expected to be as exhaustive as possible to make the review available to a large panel of users and to help the ministry to choose the “best” indicators for the assessment of its policy (Devillers et al., 2005).

### 2.2.2. Method of comparison

Each indicator was presented in a descriptive sheet, with a list of 25 criteria, a short presentation of the calculation, and the list of the parameters and variables used for calculation (Devillers et al., 2005). The following criteria were used (1) some general descriptors on the use, users and planned use, (2) others on the spatial scale, the environmental compartments taken into account and the calculation method, (3) some information useful for assessing the qualities of the indicators, the calculation time, and the existence of a scientific validation procedure according to the framework of Bockstaller and Girardin (2003), and (4) finally, four evaluation criteria expressed on a qualitative four-level scale: -, -, +, ++ covering the readability, the feasibility, the reproducibility and the relevance for the end-users. All the indicator sheets as well as the evaluation criteria were filled in by the same person and validated by a group of experts. Information sources were the references from grey and scientific literature. No implementation test was presented in this book. For the sake of concision, the number of indicators presented in this article was reduced to a selection of indicators chosen for their illustrative value or because they are already implemented (see Tab. IV in Sect. 3.2 of the “results” chapter).

## 2.3. Comparison of 5 assessment methods of sustainability in France

### 2.3.1. Context of the work

This work was launched by a regional organisation, Agro-Transfert, at the request of the agricultural sector’s representatives to develop a quality management and environmental management approach in the Picardie region, North of France (Galan et al., 2007). The first step was to develop a regional benchmark for good farming practice “Quali’terre” (Aubry et al., 2005). The second step (developed as an extension to the “Quali’terre” programme) is the development of an environmental management system which is relevant and user-friendly. In order to have a state of the art of the existing tools and to choose the best fitted tool, Agro-Transfert performed a comparison of the five methods used most frequently in France to evaluate the environmental impacts of agriculture. Those methods are all based on a set of indicators addressing different environmental themes, whereas the first method below also includes the economic and social dimensions of sustainability. The five methods are:

- (1) IDEA (“Indicateur de Durabilité des Exploitations Agricoles”), which was developed by a working group under the patronage of the French Ministry of Agriculture (Vilain et al., 2008).
- (2) DIAGE (“DIAGnostic Global d’Exploitation”), which was developed by the Regional Federation of Agricultural Cooperatives (FRCA) in the French “Centre” administrative region, in partnership with agricultural technical institutes (FRCA Centre, 2002).
- (3) DIALECTE (“DIAGnostic Liant Environnement et CTE”), which was developed by the Solagro association (Solagro, 2000) as well as the next method.
- (4) DIALOGUE (“Diagnostic agri-environnemental global d’exploitation”), which addressed more themes than DIALECTE at field level (Solagro, 2001).
- (5) INDIGO® (“indicateurs de diagnostic global à la parcelle”), which was developed by the INRA’s Sustainable Agriculture Research Unit in Colmar (Bockstaller et al., 1997).

### 2.3.2. Method of comparison

As for previous work, a set of criteria was selected by the authors to compare the methods: (1) general criteria: type of agricultural production evaluated, spatial scales, implementation time, target users, spreading and developers; (2) environmental themes and impacts, (3) main activities, crop rotation, nitrogen fertilisation, etc., (4) aggregation levels, calculation method, rating scores and thresholds, and (5) type of data required (field data, management at farm level, sensitivity of the environment). Unlike the second case study on 43 pesticide indicators where each indicator was described and evaluated in a separate sheet, the methods are here compared directly in tables.

To get some of those data, e.g. implementation time, the authors tested the five methods on a set of 15 farms in Picardie (all with cereals and sugar beet, +450 ewes for 1 farm, +50 beef for 1 farm, + potatoes for 3 farms, + vegetables for 1 farm, size ranging between 93 and 460 ha). The results obtained with each method on the 15 farms were compared in two ways:

- For a single impact, the results for all 15 farms were compared with each of the five tools. The effect of crop protection on water quality was selected.
- For each method, the individual result for four different activities (management of inert waste, nitrogen fertilisation, crop protection and energy management) within the “water pollution” theme were compared on one particular farm.

The results of the indicators were normalised by expressing them as a percentage of the maximum possible rating for the indicator, so that they can be compared (Nardo et al., 2005).

## 2.4. Comparison of four farm management tools in the upper Rhine plain (COMETE project)

### 2.4.1. Context of the work

The last work was initiated in a transregional context, in the upper Rhine plain by French, German and Swiss partners in 2003. The French and Swiss methods were compared with two German tools widely used in Germany. As in the previous

**Table I.** List of evaluation criteria used in the COMETE project (Bockstaller et al., 2006).

Scientific soundness	Feasibility	Utility
Coverage of environmental issues	Accessibility of data <sup>a</sup>	Coverage of needs <sup>a</sup>
Coverage of agricultural production branches	Qualification of user	Clearness of conclusion from results
Coverage of production factor	Need for external support	Quality of communication of results
Indicator type <sup>b</sup> , depth of environmental analysis	User-friendliness	
Avoidance of incorrect conclusions	Integration with existing farming software	
Transparency	time requirement	

<sup>a</sup> For three user groups: farmers, advisers, administration.

<sup>b</sup> Based on the driving-force, pressure, state, impact, response framework (EEA, 2005).

study, the four selected methods based on a set of environmental indicators are:

- (1) INDIGO<sup>®</sup>, also compared in the previous project (see Sect. 2.3.2).
- (2) SALCA (“Swiss Agricultural Life Cycle Assessment), developed at the Agroscope ART Reckenholz in Zurich (Switzerland), (Rossier and Gaillard, 2004).
- (3) KUL/USL (Criteria and Standards for Sustainable Agriculture), developed at the state agricultural institute of Thuringe in Iena (Germany), (Eckert et al., 2000).
- (4) REPRO, developed at the University of Halle (Germany), (Hülsbergen, 2003).

The tools were assessed according to the version valid in mid-2004. For REPRO, only a subset of the whole indicator set with high relevance for environmental items was analysed.

#### 2.4.2. Method of comparison

Since no adapted methodological framework for comparison was found in the literature, the working group of the COMETE project developed its own approach, which consists of two stages:

- First, a comparative evaluation using a list of criteria which were grouped into three domains (“scientific soundness”, “feasibility” and “utility”) (Tab. I)
- Second, the test of the implementation of the methods in a set of 13 farms. For the first step, for each criterion, a score between 1 (the lowest) and 5 (the highest) was defined by a set of decision rules. An example is given in Table II, the details being available in Bockstaller et al. (2006). The criteria addressing the users’ needs and the whole list were discussed during a workshop with the three identified user groups: farmers, advisers and agents of administration.

The four methods were evaluated by the authors themselves for INDIGO<sup>®</sup> and SALCA and validated by the whole working group. For the German methods the authors did not take part in the project, so the group decided to send the evaluation

**Table II.** Example of assessment for the criterion: “avoidance of incorrect conclusions”.

Decision rules for the assessment of the criterion “avoidance of incorrect conclusions”	Score (1 to 5)
Lack of data on evaluation of the indicator and criteria “indicator type” = 1	1
Indicator based on a non-validated model	1
No agreement of indicator value with observed data	1
Indicator criticised in a peer-reviewed article	2
Indicator based on a partially validated model	2
Lack of data on evaluation of the indicator and criteria “indicator type” = 2 to 5	2
Medium agreement of indicator value with observed data	3
Calculation method recommended by experts	3
Scientific peer-reviewed article on the indicator	4
Indicator based on a validated model	4
Good agreement of indicator value with observed data	5

carried out by the German partner to the developers of the two methods. The feedback of the latter was validated by the working group. The previous evaluation was completed by a test of the methods on a group of 13 farms (3 in Switzerland, 5 in France and 5 in Germany) for two years. The type of production was various, arable farms (maize monoculture, cereals), arable farms with special crops and mixed farms (arable crops and cattle or milk)

Following the evaluation with a set of criteria, the results obtained on the group of farms were compared in two ways. First, an aggregated indicator was calculated by means of an average value which was weighted for SALCA according to the experience gained by sensitivity analysis by the authors, without a weighting procedure for INDIGO<sup>®</sup> and KUL, and a sum of scores for REPRO. The ranking of farms obtained with each aggregated indicator was compared by means of Spearman’s correlation coefficient. Second, the conformity of recommendation derived from the indicators was compared by

**Table III.** Comparison of “nitrogen indicators” (CORPEN, 2006).

Indicator	Spatial scale	Threshold value	Time for interpretation	Agronomic relevance	Feasibility
<b>Fertilisation</b>					
amount of applied nitrogen	Field/Farm/Region	local per crop	year	1	4
amount of available nitrogen	Field/Farm/Region	local per crop	year	1	3
Number of nitrogen applications (organic and mineral)	Field/Farm/Region	local per crop	year	1	3
Deviation from the recommendation of nitrogen rate	Field	zero	year	2	3
Period of application	Field/Farm/Region	local	year	1	3
Number of grazing days	Field/Farm/Region	local	year	3	2
<b>Soil cover</b>					
Area with bare soil during drainage period	Farm/Region	local	3–4 years	2	4
Area with catch crops	Farm/Region	local	3–4 years	2	4
Assessment of surpluses or losses					
Input/output budget (CORPEN)	Field/Farm/Region	local per cropping system	≥ 5 years	2	3
N supply / requirement budget (EQUIF)	Field/Farm/Region	local: close to zero	year	3	2
Soil mineral nitrogen at harvest <sup>a</sup>	Field	local per soil type	year	3	2
Soil mineral nitrogen at beginning winter <sup>a</sup>	Field	local per soil type	year	4	2
Model predicting N losses: I <sub>N</sub> INDIGO <sup>®</sup>	Field/Farm/Region	7 (matching a concentration below roots of 50 NO <sub>3</sub> mg/L)	year	4	3 <sup>b</sup>
Model predicting Nitrate lixiviation: DEAC	Field/Farm/Region	local per cropping system	year	4	3 <sup>b</sup>

<sup>a</sup> Measured, or assessed by a model; <sup>b</sup> when the parametrisation has been achieved.

means of a newly developed index of conformity ranging between 0 (no conformity between the methods) and 1 (total conformity between the methods):

$$I_K = 1 - \left[ \frac{\sum_{p=1-n} \sum_{q=1-b} \sum_{r=1-vk} |i_{pqr} - j_{pqr}|}{(2nb)} \right]$$

with:

$i_{pqr}$ : degree of achievement of recommendation  $r$  for the production factor  $q$  for farm  $p$  for method 1;

$j_{pqr}$ : degree of achievement of recommendation  $r$  for the production factor  $q$  for farm  $p$  for method 2;

$n, b, vk$ : respectively, number of, farms, production factors and recommendations per production factor.

For example, the production factor “nitrogen management” was decomposed into recommendations like “reduce the amount of fertiliser”; “increase the amount of fertiliser”, “change the type of fertiliser” and “change the date, method of fertilisation”. If a method gives the recommendation “reduce the amount of fertiliser”, the degree of achievement will be 1 for this recommendation and 0 for the other recommendation. It should be noticed that a value inferior to 1 can be given if more than one recommendation is given.

### 3. MAIN RESULTS OF THE FOUR CASE STUDIES

#### 3.1. Comparison of indicators assessing nitrogen losses

Several groups of indicators can be distinguished in Table III: (1) a first group of simple indicators (Bockstaller et al., 2008) focusing on nitrogen input management, mainly mineral/organic fertilisation, but also organic input due to grazing. They are considered as descriptors of practices; (2) a second group addressing soil cover in winter, assessing nitrogen uptake during the period after harvest until winter, and (3) a third group resulting from the combination of variables such as nitrogen balance or model-based. Some of them are based on calculation of the input-output balance to estimate surplus. Others include nitrogen cycle processes to estimate fluxes/emission of nitrogen. Among them, the nitrogen indicator from the INDIGO<sup>®</sup> method ( $I_N$ ), based on an operational model, provides the amount of nitrogen lost to water and air (Bockstaller et al., 2008), whereas DEAC focuses on nitrate leaching in winter (Cariolle, 2002). The evaluation of the relevance and feasibility shows a relative discrepancy between the feasibility and relevance for the first and the last group in Table III. Indicators from the first group are straightforward to

calculate (high feasibility) but not really relevant if they are used alone. In contrast, indicators including in their equation nitrogen cycle processes gain in relevancy to the detriment of feasibility. In the description sheet of each indicator, recommendations are given to the users about interpretation of results and the domain of validity, and propositions of complementary indicators are given to improve the relevance of the first group. An example can be given for indicators based on the calculation of a balance (input minus output) used by several authors and institutions as an indicator for nitrogen losses (e.g. Goodlass et al., 2003; EEA, 2005). However, several authors (Lord et al., 2002; Oenema et al., 2005; ten Berge et al., 2007) pointed out by comparison with measurements of nitrate leaching that such nitrogen balances are bad estimators of nitrate leaching risk, if they are used on an annual basis (Laurent et al., 2000). Thus, the report recommended an interpretation based on pluriannual calculation.

### 3.2. Comparison of 43 pesticide risk indicators

The output of the work was a book describing the 43 indicators, 24 in a detailed way and 19 in a simplified way. Several groups of indicators can be distinguished: (1) indicators resulting from transformation of variables into scores and summed up or aggregated in an empirical way, among them EIQ, one of the first indicators published (Tab. IV); (2) a second group of indicators uses outputs from model calculation. 14 indicators among the 43 are based on the risk ratio approach which is used in registration of pesticides (Vercruyse and Steurbaut, 2002): it is the quotient of the estimated human exposure or predicted concentration and toxicological reference value used for different environmental compartments, e.g. EPRIP, POCER. (3) The third group contains specific approaches such as the qualitative one based on decision rules associated with fuzzy logic (e.g. I-Phy) or based on a multicriteria ranking method (Vaillant et al., 1995; Aurousseau, 2004).

Other trends which can be pointed out through this comparison is the lack of indicators which were validated by comparison with experimental data (12 among the 43), only one (EYP) being validated by end-users (Bockstaller and Girardin, 2003). Most of the indicators are calculated on the field scale and only 3 among 43 on the watershed scale, which is relevant for assessment of surface water quality. The implementation of the indicator requires in general less than 1 hour per calculation, except for EPRIP and EYP, which need more time because of the high number of data for calculation. Only 8 among 43 propose reference values which help users in the interpretation of the outputs. No specific focus was put on the use of the indicators.

### 3.3. Comparison of 5 assessment methods of sustainability in France

The first part of the work is descriptive. A synthesis of the results is given in Table V. Besides general information, Galan

et al. (2007) assess on a qualitative scale the degree of coverage of environmental themes and farm activities (practices) at field as well as farm level. For the first item, “water quality (sporadic pollution)”, “air quality” and “social environment (noise, odours)” are not covered by a majority of methods, whereas for the second item, most of the methods neglect or poorly integrate the activities “construction/modification of buildings or storage”, “production of renewable energy” and “management of inert waste”. Additional information is given on the type of data needed for which INDIGO® differentiates from the others by using detailed field practice data and data on the sensitivity of the environment, soil and climate, but no site data such as maintenance of the storage tank or sprayer, or building management. About the aggregation of the information, most of the methods use a simple method based on the sum of scores, and product (for DIAGE), whereas indicators in INDIGO® are based on models and expert systems (Bockstaller et al., 2008).

The authors go a step further by comparing the five assessment methods for water quality. They compare the impact of pesticide use on 15 farms. The normalised values of the pesticide indicators are represented in Figure 1. All the methods except DIAGE, and DIALOGUE to a lesser extent, show significant variations between farms. IDEA yields in general higher results, showing less impact on water quality, than the other methods. In any case, no correlation between methods appears on the sample of farms, which means that the recommendations for pesticide management will not be the same between methods for a given farm. This can be explained by the difference between methods in: (1) the integration of aspects of sporadic pollution (point source), as is the case for IDEA and DIAGE, (2) type of data used, pesticide properties (INDIGO® and DIALOGUE), and soil and environment sensitivity (INDIGO® and DIAGE), and (3) the aggregation method. Similar discrepancies between the five methods are found for one particular farm when they are compared on four different activities (management of inert waste, nitrogen fertilisation, crop protection and energy management).

### 3.4. Comparison of four farm management tools in the upper Rhine plain (COMETE project)

Based on the versions available in mid-2004 for the four methods and on a subset of indicators for REPRO, the results yielded by each method for the 15 criteria are shown in Figure 2. For the domain “scientific soundness”, SALCA presents the best environmental scores, but none of the methods was able to cover all relevant environmental issues, especially regarding biodiversity. The low scores of INDIGO® for the criteria “coverage of agricultural production” and “consideration of production factors” result from its specialisation in plant production. However, this method allows a detailed analysis of a cropping system, enabling the user to trace the cause of an environmental risk to the management, e.g. risk analysis of each pesticide application, taking into account the field conditions, tillage, spraying techniques and active ingredient properties. The “depth of environmental analysis” is low for REPRO due

**Table IV.** Examples from the comparison of 43 pesticide risk indicators (Devillers et al., 2005).

Indicator	Developer/ reference	Target user <sup>b</sup>	Spatial scale	Environ. compartment addressed <sup>a</sup>	Calculation method	Validation (D, O, U) <sup>c</sup>	Time for data (for calculation)	Readability <sup>d</sup>	Feasibility <sup>d</sup>	Reproducibility <sup>d</sup>	Relevance for user
ADSCOR	OCDE group	DM	Field/region/ country	Sw	Sum of scores	D, O	< 1 h (< 1 h)	-	+	++	+
EIQ	Kovach et al., 1992	DM, R, T, F	Field	Sw, Gw, Bio, Hf, Hp	Some/ product of scores	D, O	< 1 h (< 1 h)	-	+	++	+
EPRIP	Trevisan et al. in Reus et al., 1999	DM, R, T, F	Field	Sw, Gw, Air, Soil	Risk ratio	D, O	> 1 h (< 1 h)	+	+	+	+
EYP	Reus and Leenderste, 2000	DM, F, other	Field/farm/ region/country	Sw, Gw, Soil	Risk ratio	D, U	< 1 h (< 1 h)	+	+	++	++
I-Phy	Van der Werf and Zimmer, 1998	DM, R, T, F	Field/farm	Sw, Gw, Air,	decision tree + fuzzy logic	D, O	< 1 h (< 1 h)	++	+	++	+
p-EMA	Hart et al., 2003	D, T, F	Field/farm	Sw, Gw, Soil, Bio, Hf, Hp	Risk ratio	D	> 1 h (< 1 h)	++	++	+	++
PERI	Nilsson 1999 in Reus et al., 1999	T, F	Field/farm	Sw, Gw, Air, Soil	Some/ product of scores	-	< 1 h (< 1 h)	+	++	++	+
POCER	Vercruyse and Steurbaut, 2002	DM, R, F	Field/farm	Sw, Gw, Soil, Bio, Hf, Hp	Risk ratio	D	< 1 h (< 1 h)	++	+	+	+
REXTOX	OCDE group	DM	Field/farm/ region/country	Sw	Risk ratio	D	< 1 h (< 1 h)	-	+	++	+
SIRIS	Vaillant et al., 1995	DM, R, T	Region/country	Sw,	Ranking method	D	< 1 h (< 1 h)	-	+	++	+
SYNOPS	Gütsche and Rossberg in Reus et al., 1999	DM, T, F	Field/farm/ region/country	Sw, Soil	Risk ratio	D	< 1 h (< 1 h)	-	+	++	+
SyPEP	Pussemier in Reuset al., 2002	DM, R, F	Water catchment	Sw, Gw,	Risk ratio	-	< 1 h (< 1 h)	+	+	++	++

<sup>a</sup> Sw: surface water, Gw: groundwater, Bio: biodiversity, Hf: health of farmer, Hp: health of population, other: other compartment.

<sup>b</sup> DM: decision-maker, R: researchers, T: technicians, F: farmers.

<sup>c</sup> D: validation of the design (peer-reviewed article or by experts), O: validation of output with experimental data, U: end-user validation.

<sup>d</sup> On a four-class scale: - (low), -, +, ++ (high).

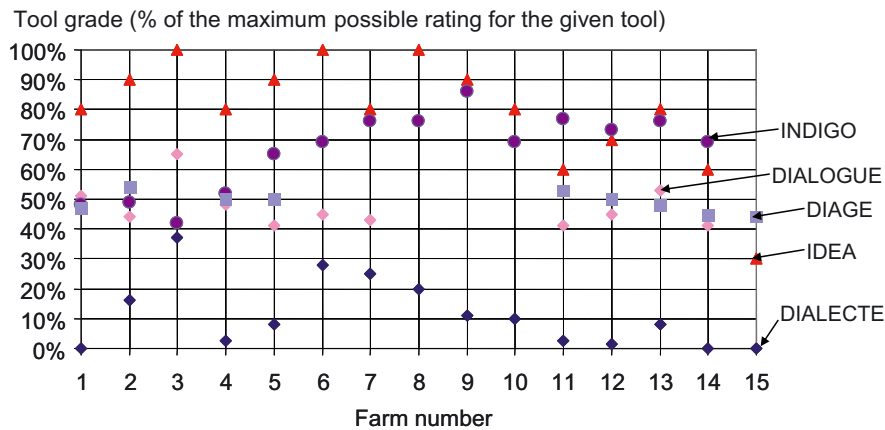


Table V. Comparison of 5 assessment methods of environmental sustainability in France (Galan et al., 2007).

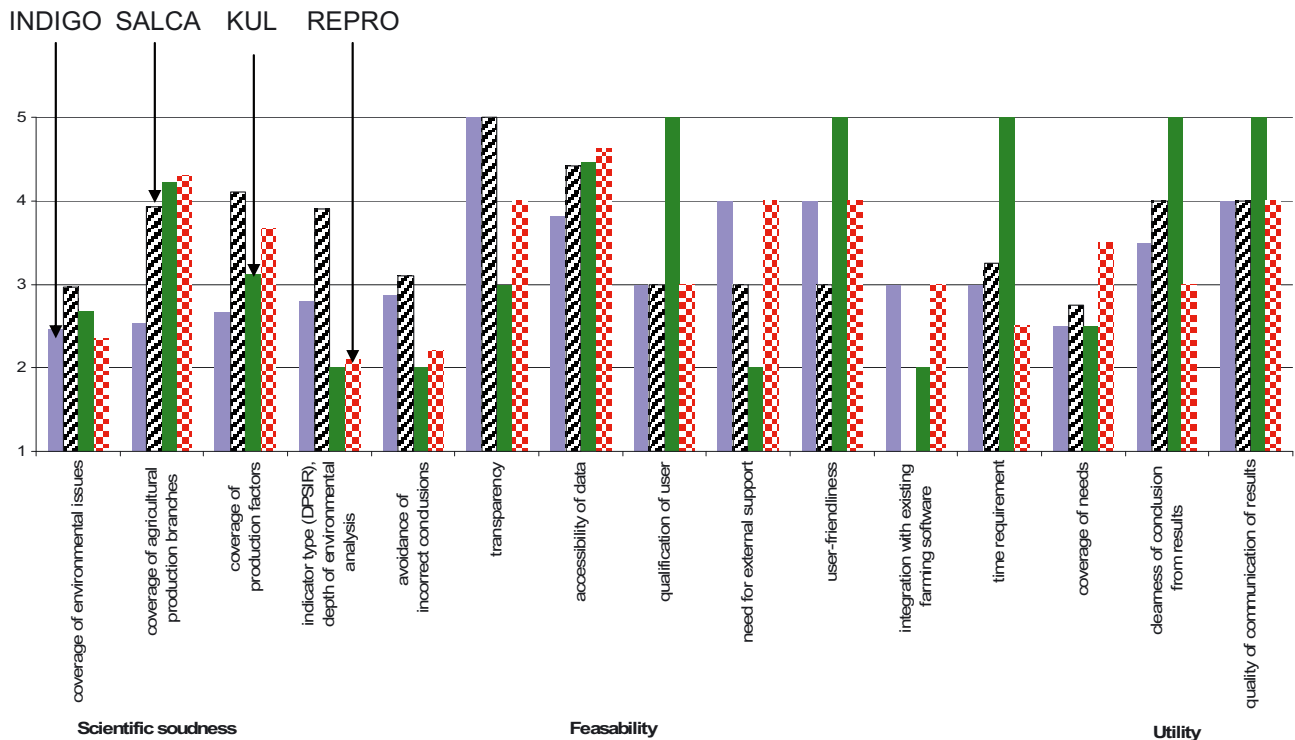
Method	Type of production	Spatial scale	Impact not or poorly addressed <sup>a</sup>	Activity not or poorly tackled <sup>b</sup>	Type of data	Indicator type	Aggregation method	Implementation time	Target users
IDEA	Crops, Animal husbandry, Market gardening	Farm	Social environment	P, K fertilisation Buildings or storage	Field practices (mean data) Site practices	Simple, composite, system indicators.	Sum	1 day	Farmer, Student, Technician, Teacher.
DIAGE	Crops, Animal husbandry, viticulture	Farm	Air Quality Natural Resources Use of non-renewable resources	Cropping pattern Crop rotation Renewable energy Management of inert waste	Field practices (mean data) Site practices Sensitivity to the environment	Composite indicators	Product-sum	1.5 days	Farmer, Technician,
DIALECTE	Crops, Animal husbandry,	Farm	Water Quality (sporadic pollution) Soil Chemical quality Air Quality Social environment	Crop protection Nitrogen fertilisation P, K fertilisation Organic manure Water management, Buildings or storage, Renewable energy Management of inert waste	Field practices (mean data) Site practices	Simple, composite, system indicators	Sum,	1–1.5 days	Farmer, Technician,
DIALOGUE	Crops, Animal husbandry,	Field, farm	Water Quality (sporadic pollution) Soil Chemical quality Air Quality Social environment	N fertilisation P, K fertilisation Organic manure Renewable energy, Management of inert waste	Field practices (per field) Site practices	Simple, composite, system indicators	Sum	2 days	Farmer, Technician,
INDIGO <sup>®</sup>	Crops, (grassland) viticulture, fruit production)	Field, farm	Water Quality (sporadic pollution) Landscape Social environment	Installation of “non-productive” elements Buildings or storage Renewable energy Management of inert waste	Field practices (per field) Sensitivity to the environment	Composite indicators	Expert system, model	2 days	Farmer, Technician, Researcher

<sup>a</sup> In Galan et al. (2007) positive list with: Water Quality (diffuse pollution), Water Quality (sporadic pollution), Soil Physical quality (structure, loss), Soil Chemical quality (chemical pollution), Air Quality, Landscape Fit with the landscape Natural Milieus Domestic and wild biodiversity, Natural Resources Use of non-renewable resources, Natural Resources Use of water reserves, Social environment (noise pollution, odours, etc.).

<sup>b</sup> In Galan et al. (2007) positive list with: Crop protection, N fertilisation, P K fertilisation, Organic manure, Soil management, Servicing and maintenance of machinery, Cropping pattern, Crop rotation, Installation of “non-productive” elements, Water management, Construction/modification of buildings or storage, Production of renewable energy, Management of inert waste.



**Figure 1.** Comparison of the output of pesticide indicators for water quality from five assessment methods. Indicators are calculated on 15 farms and their outputs are normalised as a percentage of the maximum value (Galan et al., 2007).

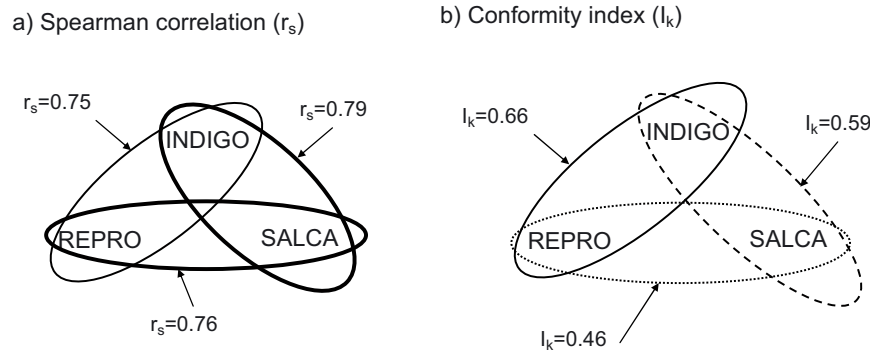


**Figure 2.** Comparison of four farm management tools in the upper Rhine plain with the help of 15 criteria (see Tab. I) in the frame of the COMETE project (Bockstaller et al., 2006).

to the fact that this method considers for each environmental issue all types of indicators without priority despite the risk of redundancy between them; and for KUL, due to the type of indicator (mainly only driving forces). Those take into account only farmers’ practices and not emissions or impacts. The low score of KUL/USL for the criterion “transparency” reflects the non-accessibility of the software, which is balanced by the score in the domain “feasibility” for which KUL/USL receives the best score as a result of its cleverly devised organisation form. On the contrary, SALCA’s electronic entry data form was not user-friendly. The evaluation with REPRO is

comparatively more time-consuming. For the domain “utility”, no great differences were observed between the four methods. The better score of KUL/USL is due to the criterion “communicability” thanks to the possibility of labelling, which is compensated for by the lack of specific recommendations at field level.

There was a high correlation between SALCA, REPRO and INDIGO® (not enough farms for KUL/USL) regarding the environmental ranking of the analysed farms (Spearman coefficients range between 0.72 and 0.88, see Fig. 3a). In other words, for the four methods, there is no reason to fear that



**Figure 3.** Comparison of outputs based on (a) the ranking of farms by means of the Spearman correlation coefficient ( $r_s$ ), (b) the conformity of recommendations by means of the conformity index ( $I_k$ ), (Bockstaller et al., 2006). A value 1 indicates a perfect correlation for  $r_s$  and conformity for  $I_k$ . Both comparisons are made with a sample of 13 farms with the data of 2002 (KUL not included in the comparison because it was implemented on 3 farms only).

the choice of the environmental management tool determines whether a farm performs well or badly from an environmental point of view. On the other hand, the conformity index shows a low convergence between the recommendations for the four methods (index range between 0.48 and 0.64, see Fig. 3b).

These discrepancies are explained by major conceptual differences between the investigated methods, namely: (1) in the different environmental issues considered. This can be illustrated by the phosphorus management: INDIGO® addresses soil fertility issues which can lead to a recommendation “increase the amount of fertility”, whereas SALCA focuses on eutrophication (of soil and water) and environmental soil quality aspects (here linked to heavy metals present in some fertilisers). Provided that a minimal yield is reached, SALCA does not recommend from an environmental point of view to increase the amount of fertiliser, whereas INDIGO can do it to maintain soil fertility. (2) In the production factors which are used for the calculations of indicators dealing with similar issues. INDIGO® and SALCA take into account amount of nitrogen, crop management, e.g. soil cover in winter, and soil mineralisation to assess nitrate leaching, whereas KUL and REPRO, for the indicator considered in the study, only take into account nitrogen input and output, and (3) to a lesser extent in the benchmark used to derive a recommendation for some similar indicators.

Besides the evaluation with criteria and the comparison of outputs, some general qualitative aspects were pointed out through the experience gained by implementing the method on farms. Two deserve more attention. The implementation of a method outside the country where it was developed raises several problems such as the accessibility of data or different description schemes for the same issue (especially for soil description) and bugs in the software due to national parametrisation. On the other side of the chain, the user stands alone for the interpretation of results and is not provided by any methods with an interpretation system in the software to interpret the results except for KUL. In this case, the user receives a written report with the interpretation and recommendations to improve the system. However, the user does not have access to the calculations and has to pay for those recommendations.

#### 4. DISCUSSION

In this discussion we will not discuss the results obtained for each method but focus on the methodology used to compare and evaluate assessment methods or thematic indicators. First, it should be noticed that if such a study is in many cases user-oriented, it can also help indicator or method developers to improve their methods. For example, the work on pesticide risk indicators was followed by a second project on indicator validation (Girardin et al., 2007) and on the improvement of two of them, e.g. introduction of a risk component on biodiversity. The developers of SALCA took into account the poor assessment of their method according to the criteria “integration with existing farming software” and “user-friendliness” (see Fig. 2) for the SALCA version of mid-2004. They integrated the use of commercial farm management software for the data collection and the implementation of a new software program for data validation and preparation before calculation for the last two years. The comparison of the 5 assessment methods in the third case study led the authors to develop a new method more fitted to the need of the local users.

In Table VI we synthesise the main features of the comparison and evaluation approaches used for the four case studies of this article. It highlights the variability between the approaches, explainable by a lack of a generic methodology. The criteria and their organisation vary between the case studies. Criteria on feasibility and relevance (or soundness) can be found in the four cases. This can be compared with previous studies. Hertwich et al. (1997) proposed only three criteria: “information requirement”, “tolerance for imperfect information” and “potential for undesirable outcome”. Other authors such as Gebauer and Bäuerle (2000) or Thomassen and de Boer (2005) developed a longer list organised, respectively, into different groups: “implementation” and “utility”, and, “relevance” for user, “quality” and “availability of data”. Other comparative studies remained mainly descriptive, including information on the time needed for data collection and recommendations on the type of indicators and linked issues, e.g. choice of threshold, scale of result expression (van der Werf and Petit, 2002; Halberg et al., 2005; Payraudeau and

**Table VI.** An overview of approaches used to compare indicators and assessment methods in the 4 case studies.

Authors	CORPEN (2006)	Deviller et al. (2005)	Galan et al. (2007)	Bockstaller et al. (2006)
Number of indicators/ methods compared	23 indicators	43 indicators	5 assessment methods (French)	4 assessment methods (2 German, 1 Swiss, 1 French)
Evaluation criteria	feasibility agronomic relevance	readability feasibility, reproducibility relevance for user	time for implementation environmental themes and activities tackled by the method	15 criteria grouped into : scientific soundness feasibility utility (see Tab. I)
Scale of evaluation	Semi-quantitative scoring between 1 and 4	Qualitative scale (--; -; +; ++)	Quantitative (days) Qualitative scale (-; +; ++; +++)	Semi-quantitative scoring between 1 and 5 with explicit decision rules
Implementation of the evaluation	Working group	1 person for the 43 indicators and validation by the group	Authors of the article	Authors of the method with cross- validation and feedback of method developer
Comparison table	Yes	No	Yes	Yes
Direct comparison of outputs	No	No	For a sample of 15 farms for one indicator, For one farm, and several indicators	Ranking of 13 farms for an aggregated indicator The conformity of recommendations for the sample of farms

van der Werf, 2005). Such descriptive comparison studies allow the users to know the construction methods better, and to appropriate the tools and complete the evaluation step which highlights strong and weak points of each method.

It should be noticed that the cost of implementation is not used in the four case studies or by all the authors previously quoted, although it is an important criterion (Romstad, 1999). This can be explained by the fact the studied methods were at an experimental stage, and that most costs are internalised by the method developers so that no realistic assessment of this criterion could be achieved.

From the list of criteria presented in Table VI or used by other authors, it appears that the meaning of the word can in some cases vary between authors. With regard to the feasibility, Hertwich et al. (1997), like Thomassen and de Boer (2005), linked it mainly to the availability of data, whereas it covers more aspects in the fourth case study (COMETE project), like in the work of Gebauer and Bäuerle (2000). Even within a working group like this of the CORPEN, the assessment of the criterion “relevance” was not so easy. It refers to a synthesis or even compromise of criteria such as sensitivity, representativeness, legibility and robustness, which are not so easy to specify. This explains the reason why the group of the COMETE project prefers to increase the number of criteria with the risk of providing too much information to the user. A solution to this inflation of criteria would be to synthesise the outcome of the evaluation with a multi-criteria analysis, as was proposed for social validation of indicators (Cloquell-Ballester et al., 2006).

The objective of the CORPEN group (CORPEN, 2006) was to guide the users in the selection of indicators addressing the nitrogen leaching issue in order to avoid misuse outside the domain of use, or misinterpretation. In the study of the CORPEN group, an evaluation of indicators is briefly presented in the

main text but no criteria are given in the descriptive sheet, whereas a synthesis in the form of text but no comparative tables are given in the book of Devillers et al. (2005). A database with queries to help to choose a pesticide indicator is in development (Girardin, personal communication). The third comparative case study (Galan et al., 2007) provides several tables comparing the French assessment methods for their technical features regarding their calculation method, the domain of use, etc. which could be used for an evaluation work. The time for implementation is quantified but not valued like in the last case study, the COMETE project (Bockstaller et al., 2006). The last case study, the COMETE project, clearly differentiates description and evaluation and proposes a method based on a set of criteria with decision rules to assess them (see Tab. II). This should increase the transparency. However, a degree of subjectivity may remain in the criteria of the COMETE project as some criteria are the results of a scoring procedure without decision rules, e.g. coverage of an environmental issue. The cross-validation which was done in the project could help to reduce the subjectivity. Another point to notice is the effort to make the evaluation more precise by differentiating different user groups. The authors of the COMETE project (Bockstaller et al., 2006) identify three groups (farmers, advisers and employers of administration) which are differentiated for the evaluation of two criteria, accessibility of data and coverage of needs. This was also done by Thomassen and de Boer (2005), who added a fourth group of scientists to the three groups for one criterion, “comprehensibility”. A criterion such as accessibility of data also has to be adapted to the context of use. Some data, such as those describing soils, vary a lot between countries or even regions (Bockstaller et al., 2006).

An interesting output of the third case study (Galan et al., 2007) is the comparison of the outputs of the methods, which is rarely done according to our knowledge. Examples can be

found in the literature on comparison of outputs for pesticide risk indicators (Maud et al., 2001; Reus et al., 2002). However, those authors compared the ranking of pesticides but did not take into account the absolute value of the indicator, so that the actual difference between the results of two indicators is not assessed. In the work of Galan et al., assessment methods based on different sets of indicators are compared. Consequently, Galan et al. (2007) restricted the analysis to comparisons farm by farm or indicator by indicator. In the COMETE project, results of the individual indicators are aggregated although the developers (except for REPRO) do not propose it for users because of methodological problems due, for example, to the addition of scores (Schärlig, 1985). The second approach based on a conformity index is original and avoids this problem. However, it requires an effort of formalisation of the potential recommendations for each indicator within an evaluation method. Comparisons of outputs in Galan et al. (2007), like the comparison of recommendations in COMETE, yielded poor convergence between the compared methods, which can be explained by the ground difference in assumptions and choices in the calculation methods. The potential users should be aware of this, which is only possible if those assumptions are transparent.

## 5. CONCLUSION

This article highlights through the four case studies the variability in approaches used to compare indicators or assessment methods. The first two studies focus on, respectively, 23 and 43 indicators addressing the nitrate leaching issue and pesticide risk, respectively. Those studies provide a lot of descriptive information about the indicators summarised in the article. Few evaluation criteria are used to point out strong and weak points of those indicators. The third and fourth studies compare environmental assessment methods based on indicators, respectively, five used in France and four tested in the upper Rhine plain (France, Germany and Switzerland, COMETE project). Both studies also compare the outputs of the methods and highlight a low degree of convergence among them. The approach developed in the COMETE project appears to be the most elaborate. It should be tested in other comparative studies like the third case study. An adaptation to the comparison of pesticide risk indicators is ongoing in the Endure network (Kägi et al., 2008)

Our study can contribute to developing a “meta-method” which should help with the selection of indicators or of assessment methods. Such a “meta-method” could rest on a list of criteria like those of COMETE which would require local adaptation: which criteria are relevant for a given context, but also how they should be assessed, e.g. availability of soil data, which can change between countries or even regions. It should include descriptive information, evaluation criteria based not only on theoretical information but also on a test in practice. Basic assumptions, the potentialities of the methods, e.g. environmental issues covered, factors addressed, should in any case be stated clearly because they strongly influence the final results and explain the divergence between methods in terms

of recommendations. Further work is needed to help users to cope with those potential discrepancies between indicators for the same issue, or between assessment methods.

**Acknowledgements:** The comparison of 43 pesticide indicators was sponsored by the French Ministry for Ecology and Sustainable Development. The comparison work in Picardie received financial support from the Picardie administrative region (“Conseil Régional de Picardie”) and the French government’s ADEME agency (“Agence pour l’Environnement et la Maîtrise de l’Énergie”). The research within the COMETE project was partly funded by the ITADA (Institut transfrontalier d’application et de développement agronomique) which is sponsored by the EU (Programme INTERREG 3) as well as the Swiss Confederation and the cantons of Basel-Stadt, Basel-Land and Aargau.

## REFERENCES

- Aubry C., Galan M.B., Mazé A. (2005) Garanties de qualité dans les exploitations agricoles: exemples de l’élaboration du référentiel Quali’Terre® en Picardie, *Cah. Agric.* 14, 313–321.
- Aurousseau P. (2004) Agrégation des paramètres et bases mathématiques de combinatoire de facteurs de risque, in: Barriuso E. (Ed.), *Estimation des risques environnementaux des pesticides*, INRA Editions, Paris, pp. 58–74.
- Bockstaller C., Girardin P. (2003) How to validate environmental indicators, *Agr. Syst.* 76, 639–653.
- Bockstaller C., Girardin P., Van der Werf H.G.M. (1997) Use of agro-ecological indicators for the evaluation of farming systems, *Eur. J. Agron.* 7, 261–270.
- Bockstaller C., Gaillard G., Baumgartner D., Freiermuth Knuchel R., Reinsch M., Brauner R., Unterseher E. (2006) Méthodes d’évaluation agri-environnementale des exploitations agricoles : Comparaison des méthodes INDIGO, KUL/USL, REPRO et SALCA, ITADA, Colmar, p. 112.
- Bockstaller C., Guichard L., Makowski D., Aveline A., Girardin P., Plantureux S. (2008) Agri-environmental indicators to assess cropping and farming systems. A review, *Agron. Sustain. Dev.* 28, 139–149.
- Cariolle M. (2002) DEAC – Nitrogen: means to diagnose nitrogen leaching on a mixedfarm scale, *Proceedings of the 65th Institut International de Recherches Bettervières Congress*, Brussels, Belgium, 13–14 February 2002, pp. 67–74.
- Cloquell-Ballester V.A., Monterde-Diaz R., Santamarina-Siurana M.C. (2006) Indicators validation for the improvement of environmental and social impact quantitative assessment, *Environ. Impact Assess. Rev.* 26, 79–105.
- CORPEN (2006) Des indicateurs AZOTE pour gérer des actions de maîtrise des pollutions à l’échelle de la parcelle, de l’exploitation et du territoire, Ministère de l’Écologie et du Développement Durable, [http://www.ecologie.gouv.fr/IMG/pdf/maquette\\_azote29\\_09.pdf](http://www.ecologie.gouv.fr/IMG/pdf/maquette_azote29_09.pdf), Paris, p. 113.
- Devillers J., Farret R., Girardin P., Rivière J.-L., Soulas G. (2005) Indicateurs pour évaluer les risques liés à l’utilisation des pesticides, Lavoisier, Londres, Paris, New-York.
- Eckert H., Breitschuh G., Sauerbeck D. (2000) Criteria and Standards for Sustainable Agriculture, *J. Plant Nutr. Soil Sci.* 163, 337–351.
- EEA (2005) Agriculture and environment in EU-15; the IRENA indicator report, European Environmental Agency (EEA), Copenhagen (Denemark), p. 128.
- FRCA Centre (2002) DIAGE, manuel d’utilisation et logiciel.
- Galan M.B., Peschard D., Boizard H. (2007) ISO 14 001 at the farm level: Analysis of five methods for evaluating the environmental impact of agricultural practices, *J. Environ. Manage.* 82, 341–352.

- Gebauer J., Bäuerle A.S. (2000) Betriebliche Umweltinformationstechniken für die Landwirtschaft, Ber. Landwirtsch. 78, 354–392.
- Girardin P., Devillers J., Thybaud E., Soulas G. (2007) Programmes "Indicateurs et pesticides" Phase II : Validation et proposition d'amélioration d'indicateurs "pesticides", Ministère de l'Écologie et du Développement Durable, p. 71.
- Goodlass G., Halberg N., Verschuur G. (2003) Input output accounting systems in the European community – an appraisal of their usefulness in raising awareness of environmental problems, Eur. J. Agron. 20, 17–24.
- Halberg N., van der Werf H.M.G., Basset-Mens C., Dalgaard R., de Boer I.J.M. (2005) Environmental assessment tools for the evaluation and improvement of European livestock production systems, Livest. Prod. Sci. 96, 33–50.
- Hart A., Brown C.D., Lewis K.A., Tzilivakis J. (2003) p-EMA (II): evaluating ecological risks of pesticides for a farm-level risk assessment system, Agronomie 23, 75–84.
- Hertwich E.G., Pease W.S., Koshland C.P. (1997) Evaluating the environmental impact of products and production processes: A comparison of six methods, Sci. Total Environ. 196, 13–29.
- Hülsbergen K.J. (2003) Entwicklung und Anwendung eines Bilanzierungsmodells zur Bewertung der Nachhaltigkeit landwirtschaftlicher Systeme, Shaker Verlag (Halle, Univ., Habil.-Schr., 2002), Aachen.
- Kägi T., Bockstaller C., Gaillard G., Hayer F., Mamy L., Strassmeyer J. (2008) Multicriteria evaluation of RA and LCA assessment methods considering pesticide application, European Network for Durable Exploitation of crop protection strategies (ENDURE), Internal report, p. 43, <http://www.endure-network.eu/>.
- Kirchmann H., Thorvaldsson G. (2000) Challenging targets for future agriculture, Eur. J. Agron. 20, 12, 145–161.
- Kovach J., Petzoldt C., Degni J., Tette J. (1992) A method to measure the environmental impact of pesticides, New York's Food and Life Sciences Bulletin, 8 p.
- Laurent F., Vertès F., Farrugia A., Kerveillant P. (2000) Effets de la conduite de la prairie pâturée sur la lixiviation du nitrate. Proposition pour une maîtrise du risque à la parcelle, Fourrages 164, 397–420.
- López-Ridaura S., van Keulen H., van Ittersum M.K., Leffelaar P.A. (2005) Multi-scale methodological framework to derive criteria and indicators for sustainability evaluation of peasant natural resource management systems, Environ. Dev. Sustain. 7, 51–69.
- Lord E.I., Anthony S.G., Goodlass G. (2002) Agricultural nitrogen balance and water quality in the UK, Soil Use Manage. 18, 363–369.
- Maud J., Edwards-Jones G., Quin F. (2001) Comparative evaluation of pesticide risk indices for policy development and assessment in the United Kingdom, Agr. Ecosyst. Environ. 86, 59–73.
- Meyer-Aurich A. (2005) Economic and environmental analysis of sustainable farming practices – a Bavarian case study, Agr. Syst. 86, 190–206.
- Meynard J.M., Cerf M., Guichard L., Jeuffroy M.H., Makowski D. (2002) Which decision support tools for the environmental management of nitrogen? Agronomie 22, 817–829.
- Mitchell G., May A., Mc Donald A. (1995) PICABUE: a methodological framework for the development of indicators of sustainable development, Int. J. Sust. Dev. World 2, 104–123.
- Nardo M., Saisana M., Saltelli A., Tarantola S. (2005) Tools for composite indicators building. Joint Research Center, European Commission, Ispra (Italy), p. 134.
- Nemecek Th., Huguenin-Elie O., Dubois D., Gaillard G. (2005) Ökobilanzierung von Anbausystemen im schweizerischen Acker- und Futterbau, Schriftenreihe der FAL 58, Agroscope FAL Reckenholz, 155 p., Zurich.
- Oenema O., van Liere L., Schoumans O. (2005) Effects of lowering nitrogen and phosphorus surpluses in agriculture on the quality of groundwater and surface water in the Netherlands, J. Hydrol. 304, 289–301.
- Payraudeau S., van der Werf H.M.G. (2005) Environmental impact assessment for a farming region: a review of methods, Agr. Ecosyst. Environ. 107, 1–19.
- Reus J., Leendertse P.C. (2000) The environmental yardstick for pesticides: a practical indicator used in the Netherlands, Crop Prot. 19, 637–641.
- Reus J., Leenderste P., Bockstaller C., Fomsgaard I., Gutsche V., Lewis K., Nilsson C., Pussemier L., Trevisan M., van der Werf H., Alfarroba F., Blümel S., Isart J., McGrath D., Seppälä T. (1999) Comparing environmental risk indicators for pesticides. Results of the European CAPER project. Centre for Agriculture and Environment, Utrecht, p. 183.
- Reus J., Leenderste P., Bockstaller C., Fomsgaard I., Gutsche V., Lewis K., Nilsson C., Pussemier L., Trevisan M., van der Werf H., Alfarroba F., Blümel S., Isart J., McGrath D., Seppälä T. (2002) Comparing and evaluating eight pesticide environmental risk indicators developed in Europe and recommendations for future use, Agr. Ecosyst. Environ. 90, 177–187.
- Riley J. (2001a) The indicator explosion: local needs and international challenges, Agr. Ecosyst. Environ. 87, 119–120.
- Riley J. (2001b) Indicator quality for assessment of impact of multidisciplinary systems, Agr. Ecosyst. Environ. 87, 121–128.
- Romstad E. (1999) Theoretical considerations in the development of environmental indicators, in: Brouwer F.M., Crabtree J.R. (Eds.), Environmental indicators and agricultural policy, CAB International, Wallingford, pp. 13–23.
- Rosnoblet J., Girardin P., Weinzaepflen E., Bockstaller C. (2006) Analysis of 15 years of agriculture sustainability evaluation methods, in: Fotyma M., Kaminska B. (Eds.), 9th ESA Congress, Warsaw, Poland, pp. 707–708.
- Rossier D., Gaillard G. (2004) Ökobilanzierung des Landwirtschaftsbetriebs: Méthode und Anwendung in 50 Landwirtschaftsbetrieben, 53, p. 142.
- Schärlig A. (1985) Décider sur plusieurs critères. Panorama de l'aide à la décision multicritère, Presses polytechniques et universitaires romandes, Lausanne.
- Solagro (2000) DIALECTE, Diagnostic Liant Environnement et Contrat Territorial d'Exploitation ; manuel d'utilisation et logiciel.
- Solagro (2001) DIALOGUE, Diagnostic Agri-environnemental Global d'exploitation agricole ; manuel et logiciel.
- ten Berge H.F.M., Burgers S.L.G.E., Van der Meer H.G., Schröder J.J., Van der Schoot J.R., Van Dijk J.R. (2007) Residual inorganic soil nitrogen in grass and maize on sandy soil, Environ. Pollut. 145, 22–30.
- Thomassen M.A., de Boer I.J.M. (2005) Evaluation of indicators to assess the environmental impact of dairy production systems, Agr. Ecosyst. Environ. 111, 185–199.
- Thompson M.A. (1990) Determining Impact significance in EIA: a review of 24 methodologies, J. Environ. Manage. 30, 235–250.
- Vaillant M., Jouany J., Devillers J. (1995) A multicriteria estimation of the environmental risk of chemicals with the SIRIS method, Toxicol. Model. 1, 57–72.
- van der Werf H.G.M., Petit J. (2002) Evaluation of environmental impact of agro-culture at the farm level: a comparison and analysis of 12 indicator-based methods, Agr. Ecosyst. Environ. 93, 131–145.
- van der Werf H.M.G., Zimmer C. (1998) An indicator of pesticide environmental impact based on a fuzzy expert system, Chemosphere 36, 2225–2249.
- Vercruyse F., Steurbaut W. (2002) POCER, the pesticide occupational and environmental risk indicator, Crop Prot. 21, 307–315.
- Vilain L., Boisset K., Girardin P., Guillaumin A., Mouchet C., Viaux P., Zahm F. (2008) La méthode IDEA : indicateur de durabilité des exploitations agricoles : guide d'utilisation, 3<sup>e</sup> édition, Educagri, Dijon.