



## **Agri-environmental indicators to assess cropping and farming systems. A review**

Christian Bockstaller, Laurence Guichard, David Makowski, Anne Aveline,  
Philippe Girardin, Sylvain Plantureux

### **► To cite this version:**

Christian Bockstaller, Laurence Guichard, David Makowski, Anne Aveline, Philippe Girardin, et al.. Agri-environmental indicators to assess cropping and farming systems. A review. *Agronomy for Sustainable Development*, 2008, 28 (1), pp.139-149. 10.1051/agro:2007052 . hal-00886428

**HAL Id: hal-00886428**

**<https://hal.science/hal-00886428>**

Submitted on 11 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.

# Agri-environmental indicators to assess cropping and farming systems. A review

Christian BOCKSTALLER<sup>1\*</sup>, Laurence GUICHARD<sup>2</sup>, David MAKOWSKI<sup>2</sup>, Anne AVELINE<sup>3</sup>, Philippe GIRARDIN<sup>1</sup>,  
Sylvain PLANTUREUX<sup>4</sup>

<sup>1</sup> INRA, UMR1121 INPL/ENSAIA/INRA, BP 20507, 68021 Colmar, France

<sup>2</sup> INRA, UMR211 INRA/AgroParisTech, BP 01, 78850 Thiverval-Grignon, France

<sup>3</sup> Laboratoire d'Écophysiologie Végétale et Agronomie, ESA, BP 748, 49007 Angers, France

<sup>4</sup> INPL/ENSAIA, UMR1121 INPL/ENSAIA/INRA, 2 av. Forêt de Haye, 54500 Vandœuvre, France

(Accepted 26 October 2007)

**Abstract** – Environmental impacts of agriculture cannot be always assessed by using direct measurements. Since the 1990s, numerous agri-environmental indicators were developed to assess the adverse effects of cropping and farming systems in the environment, such as water pollution, soil erosion, and emission of greenhouse gases. Here we present the different types of indicators developed during the last decade and review the progress of the methods used for their development. The application of different groups of indicators is discussed and illustrated by examples in the fields of nitrogen losses and pesticide risk: (1) indicators based on a single or a combination of variables related to farmer practices, (2) indicators derived from operational or more complex simulation models assessing emissions of pollutants, and (3) measured indicators linked directly to environmental impacts. The nitrogen indicator (IN) of the INDIGO method and the MERLIN indicator will be presented and used to illustrate the methodological discussion. We show that a good identification of the end-users, of the practical objectives of the indicator, and of the spatial and temporal scales is essential and should be done at a preliminary step before designing the indicator itself. The possibilities of deriving an indicator from a model and of setting a reference value are discussed. Several methods are also presented to study the sensitivity and the validity of agri-environmental indicators. Finally, several practical recommendations are made. As only few data are usually available at the regional level, several simple indicators should be used for assessing a given impact at this level. When more detailed information is available, indicators based on operational models can be useful to analyse the effects of several factors related to soil, climate, and cropping system on an environmental impact. In experimental studies, we suggest using both measured indicators and model-based indicators.

environmental assessment / indicators / simulation model / validation / nitrogen / pesticide

## 1. INTRODUCTION

Direct measurements of impact due to agriculture are often difficult to implement. Since the 90s, numerous agri-environmental indicators and indicator-based methods were developed to assess environmental impacts of agriculture and the sustainability of agricultural systems (Rigby et al., 2001; Rosnoblet et al., 2006). Riley (2001a) spoke about an “indicator explosion”, which could be explained by the growing concern for environmental issues and sustainability. The use of indicators constitutes an alternative to direct impact measurement (Mitchell et al., 1995) which presents several methodological difficulties such as impossibility of measurement and complexity of the system or practical constraints, e.g. costs and time. According to Gras et al. (1989): “*Indicator is a variable which supplies information on other variables which are difficult to access*”.

Assessment methods based on a set of indicators have been developed at national or international levels (e.g. EU, Delbaere and Serradilla, 2004) but, also, at regional (Payraudeau and

van der Werf, 2005), farms (Eckert et al., 2000; van der Werf and Petit, 2002; Hülsbergen, 2003; Meyer-Aurich, 2005), or, field and cropping system levels (Bockstaller et al., 1997; López-Ridaura et al., 2005). Studies on specific thematic indicators are also available, for examples for nutrients (ten Berge et al., 2002; Goodlass et al., 2003), or for pesticides (Maud et al., 2001; Reus et al., 2002; Devillers et al., 2005). The generic term of “indicator” represents a large diversity of tools and needs some clarification (Riley, 2001b).

Investigations on methodological issues regarding the development of indicators were also published. Some authors have focused on specific questions like the selection of indicators (Mitchell et al., 1995), their aggregation (Nardo et al., 2005; Jollands, 2006), and their validation (Bockstaller and Girardin, 2003; Cloquell-Ballester et al., 2006). Girardin et al. (1999) identified five steps in the development of an indicator: (1) preliminary definition of the objectives and identification of the end-users, (2) the construction of the indicator, (3) the selection of a reference value, (4) sensitivity analysis, (5) validation. All steps are not always clearly addressed in articles presenting a specific indicator or an assessment method

\* Corresponding author: bockstal@colmar.inra.fr

based on indicators. The purpose of this article is to present the diversity of existing indicators and to review methodological progress in each of the five steps defined by Girardin et al. (1999). The article focus on the agri-environmental indicators, mainly at field scale (cropping system) and farm scale (farming system) with some references to higher levels. Our concepts will be illustrated with two indicators assessing nitrogen losses.

## 2. OVERVIEW OF AGRI-ENVIRONMENTAL INDICATORS

Maurizi and Verrel (2002) present a series of definition for the notion of indicator. These definitions are strongly influenced by the background and disciplines of their authors. Authors working on the assessment of ecosystem health or biodiversity refer to a set of biophysical measurements like physico-chemical properties of soil, or abundance of species for a given taxon (Carignan and Villard, 2002; Clergué et al., 2005). Other authors working in agronomy or/and economy based their indicators on model outputs (Meyer-Aurich, 2005) whereas, for some agronomists, the use of indicators is justified by the impossibility of using dynamic models or direct field measurement (Bockstaller et al., 1997).

We based our typology on a cause-effect chain defined from the works of Hertwich et al. (1997), and, Payraudeau and van der Werf (2005), as illustrated in Figure 1. According to this conceptual framework, the environmental impact results from a chain of processes, beginning with human activity like agricultural practices which can result in emissions of pollutant, depending on the characteristics of the environment, soil, climate. If transfer conditions are favourable, those emissions may cause a change of state for a given environmental compartment. If living beings are exposed to the pollutant, a biological or economic impact can occur, depending on the behaviour of the target organism and the toxicity of the pollutant. It should be noticed that the concept of potential impact is derived from a simplification of the assessment pathway (Freyer et al., 2000). For example in some Life Cycle Analysis methods (Brentrup et al., 2004), the impact is assessed by combining emissions of pollutant and the toxicity without data on the behaviour of living beings determining exposition. Other authors use the concept of risk, especially to deal with the environmental impact of pesticides (Levitan, 2000). Many pesticide risk indicators are based on an assessment of pesticide fate and hazard (toxicity) although the term of risk in its proper meaning is a probability of occurring of a hazard and the magnitude of its effect (Flemström et al., 2004).

Several types of indicators can be distinguished in Figure 1. They are based on the cause-effect chain although exceptions can be found. (1) The first group consists of simple indicators based on the use of one type of variable obtained by survey, databases and not directly measured. In many cases, they are derived from statistics on farmers' practices which are assumed to be causes of the impact according to the knowledge of their developer. Some are also based on environment characteristics, e.g. soil, climate. They can be based on one or a

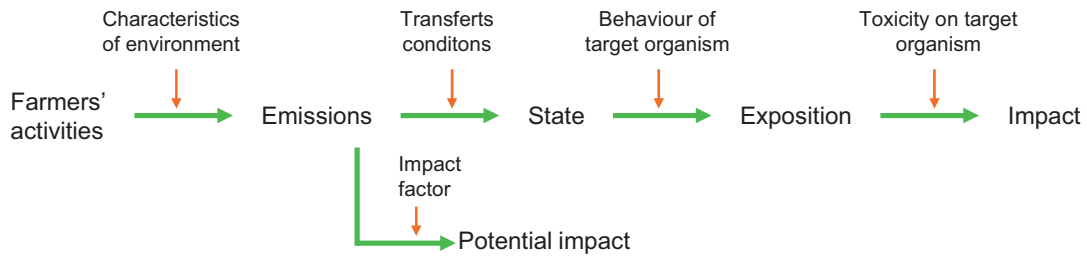
simple combination of variables, like the calculation of balances, e.g. for nutrients, or ratio, e.g. for energy. Indicators of this group provide an indirect assessment of the environmental impacts and are often qualified as "proxy", having a poor quality of prediction (Riley, 2001c). (2) The second group of indicators includes indicators based on calculation and integrating more than one type of factors, e.g. farm practices and soil conditions. This category covers a great diversity of indicators with different levels of complexity, from emission coefficient used in Life Cycle Analysis, to indicators based on the mechanistic simulation models. Those indicators are often used to assess emissions of pollutant or the pollution of an environmental component like water compartments. Advantages and limitations depend on the type of tools from which the indicator is derived. Conceptual or mechanistic simulation models integrating well processes may be preferred from the scientific point of view. They allow to link the predicted effect to causes. However their complexity is a major limitation to use in many cases. Solution can be found as shown in Section 3.2.2. (3) The third group includes indicators based on one or several measurements. Biodiversity indices belong to this category (Carignan and Villard, 2002; Clergué et al., 2005). They are used when users focus on impacts and when no accurate model is available. Emissions can also be assessed by measurements, e.g. mineral nitrogen in soil before winter, nitrate concentration below roots measured by ceramic cups. The drawbacks of these indicators are that their costs can be high and that they cannot be used to trace cause-effect relationship with a satisfactory level of accuracy (Merkle and Kaupenjohann, 2000). For example a given level of mineral nitrogen in soil can be explained by the soil and climate, the crop yield, the nitrogen management. Hence it is not easy to derive directly advices for farmers.

## 3. METHODOLOGICAL ISSUES FOR DESIGNING AGRI-ENVIRONMENTAL INDICATORS

A part of this section concerns the three groups of indicators presented above, but some specific points only concern the category 'calculated indicators' and, more specifically, indicators derived from models. This section treats the development of single indicators as well as sets of indicators.

### 3.1. Preliminary choices and assumptions

The identification of the end-users and the definition of the practical objectives of the indicator, were pointed out as an essential step by several authors (Mitchell et al., 1995; Crabtree and Brouwer, 1999; Girardin et al., 1999; Yli-Viikari et al., 2007). This preliminary step will serve as a basis to design the indicator and to evaluate its quality. Different users group can be identified like, for example, scientists, advisors, farmers, decision maker, or consumers. The group of people doing the calculations and the group of people using the results should be differentiated. In general, a given indicator will be adapted

Cause effect chainIndicator typeExamples of indicators :

Variable Single	combination	Emission factor, operational models, mechanistics models	Measurement
<b>Nitrogen</b> Rate N % soil cover in winter % non fertilized area	Farm gate N balance Surface N balance Leaching index (=drainage/ available water capacity)	NH3 emission coef. MERLIN* Nitrogen indicator IN* Output of leaching model	Soil mineral nitrogen in winter Nitrate concentration in dwell
<b>Pesticide</b> Number of treatments Amount of active ingredient	IFT (sum of rate/registered rate)	I-Phy P-ema (using output of the Macro model)	pesticide concentration in dwell, river, number of death due to pesticides

**Figure 1.** A typology of indicators based on the cause-effect chain of impact: examples of indicators, variable based indicator, indicator derived from a model, and measured indicators are presented for assessing nitrogen losses and risks induced by pesticides. Examples of indicator with asterisk will be detailed later in text.

to one group of stakeholders only due to the variability of the users' requirements.

An indicator can be developed for various objectives like *ex ante* evaluation of actions, in a planning phase (Sadok et al., 2007), *ex post* evaluation of an action at the end or during its implementation, monitoring purpose with an alert role, decision support in real time to drive the system, communication. After clarifications of those general items the developer should focus on the issues of concern covered by the indicators, e.g. the environmental compartment or impact.

Those are generally translated into a list or even more into a framework which is more or less explicit and elaborated. This list of issues can be set up in interactions with different users-groups or experts, by consultation, e.g. with Delphi techniques (Hess et al., 1999). Girardin et al. (1999) proposed to synthesize them in a matrix crossing issues of concern and element of the system to assess. Actually two groups of approaches can be distinguished according to two conception of sustainability (Hansen, 1996). The former one is a goal-oriented, based on a set of themes or objectives (von Wirén-Lehr, 2001). Those can address the main abiotic (air, soil, water) and biotic

(species, ecosystems) environmental components, ecological functions or environmental impacts like in Life Cycle Analysis (Brenttrup et al., 2004). The latter is property-oriented, based on systemic properties e.g. adaptability, security. A synthesis of attributes of properties can be found in López-Ridaura et al. (2005). Bossel (1999) showed that sustainability can be assessed by means of a set of seven generic systemic properties, such as existence, effectiveness, freedom of action, security, adaptability, coexistence, psychological needs. This systemic approach is an alternative to the goal-oriented approach which Bossel (1999) describes as based on the "intuition" of experts and contingencies. It should also help to reduce the number of indicators. However, the concept of systemic property appears to be abstract to non-initiated users. Efforts to make it operational are still needed to help the user to select relevant indicators for each property.

Last the definition of the system boundaries is another important step (Van Cauwenbergh et al., 2007). It includes the calculation scales, spatial and temporal which will be influenced by the users' needs, the issues of concerns, etc. Again the choice will guide the type of indicators and the required

qualities. For example, indicators calculated on data obtained at regional or national levels should show good statistical qualities. In Life Cycle Analysis approaches, users are forced to define the system boundaries. It can be the product, the farm including or not upstream such as production of inputs and off-stream activities such as waste management. In other assessment methods quoted in the introduction, this definition seems to be neither explicit nor unified between indicators. Regarding spatial and time scales one should paid attention to the resolution of calculation, the level at which basic calculations are carried out. Farm and year are typical resolution for environmental indicators. This should not be confused with the extent, i.e. the whole area, e.g. the region, or time span, e.g. period, crop rotation, covered by the indicators calculation (Purtauf et al., 2005).

## 3.2. Indicator design

### 3.2.1. Nature of the indicator outputs

As shown in Figure 1 indicator outputs may represent the result of a measurement, the result of a simple calculation based on a combination of data, or the result of a simulation derived from a complex model. Such outputs can be transformed into a score which expresses (1) a risk or an impact, ranging from 0 (low) to 1 (high) (van der Werf and Zimmer, 1998). Other authors used a scale between 1 and 10 (Eckert et al., 2000) or 1 and 5 (see Fig. 3), (2) an environmental performance ranging between 0 (low), and 10 as in Bockstaller et al. (1997). (3) Scales between a negative value and a positive one, e.g. -3 to +3 as in Rigby et al. (2001), expressing a negative and positive effect respectively are also used. The choices of the scale, of the scoring function and of the range of value are subjective, will depend on practical considerations, and can be subject to discussion (Andreoli et al., 1999). They have an importance for communication. In any case, these choices should be explicit and transparent.

### 3.2.2. Model-based indicators

A model output can be used to calculate an indicator. This option is attractive by the potentialities of modelling, but may lead to some practical problems due to the complexity of many models. A solution is to create a matrix of simulations and to derive an indicator from this matrix. The interest of this approach is that the model is run for a limited number of situations. For example Brown et al. (2003) used the model MACRO to build a table of concentrations of pesticides in groundwater from 2280 model simulations. Another approach is to derive a metamodel which can be then used to calculate an indicator (Garcet et al., 2006). Such a tool can be elaborated with a learning machine (Shan et al., 2006). Finally, another solution is to develop a simplified model based on a reduced number of input variables which are easily accessible. The latter approach is illustrated by the nitrogen indicator presented

below (Fig. 2). Such an operational modelling can be characterized by the statement of Durand et al. (2002): “... *to compare the effect of different agricultural practices on nitrogen pollution in a catchment, it may not be necessary to simulate quantitatively the whole nitrogen and water cycle to calculate nitrate concentration in the river at each time step*”.

### 3.2.3. Qualitative approach

An indicator can also be qualitative. It can represent a decision rule which can be expressed as a “if then” rule, or as a contingency table. This will be illustrated by the example of MERLIN presented in Section 3.4. Fuzzy logic can be used to account for the uncertainty in the indicator outputs by avoiding the effect of knife-edge limit of a given class like the ones used in MERLIN (Fig. 3), (Silvert, 2000; Enea and Salemi, 2001). It is applied in a growing number of examples, e.g. van der Werf and Zimmer (1998); Prato (2005). Ranking method like SIRIS based on scoring was also proposed to derive qualitative indicators (Aurousseau, 2004).

## 3.3. The setting of a reference value

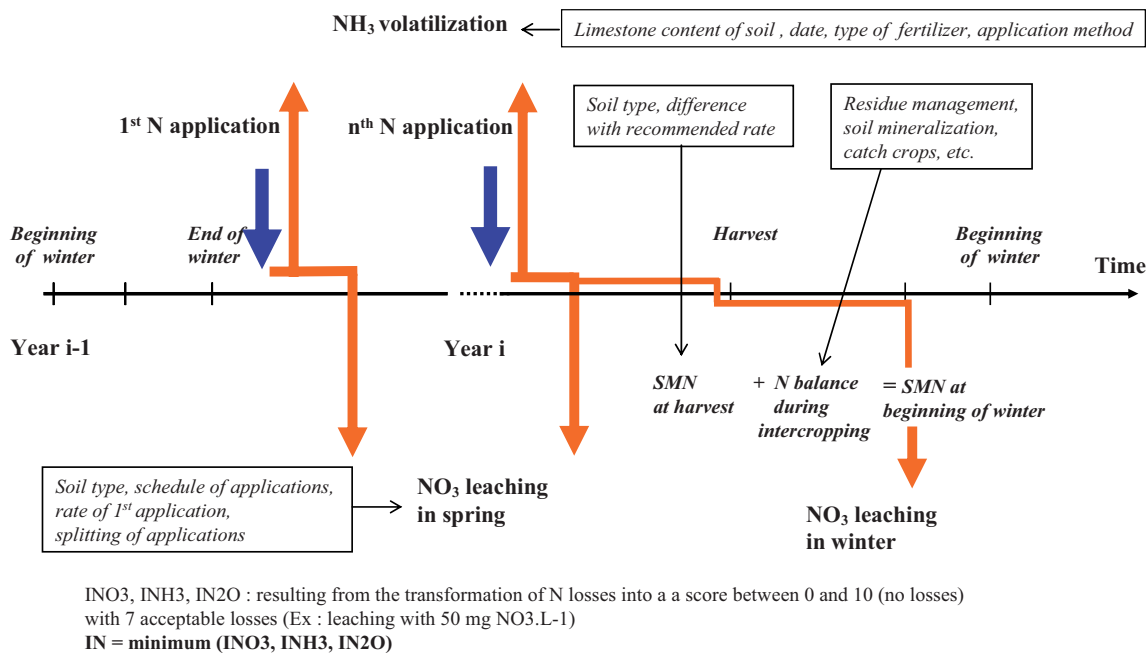
According to Riley (2001c), indicators are defined as “observations relative to their respective reference point”. This reference value helps the user to interpret the raw value of the variable, the calculation or the measurement, for instance to assess whether “an action A is environmental friendly or not”. The reference value may be implicit. For example, the reference value for nitrogen balance indicators is zero for many users, assuming that the system has reached a steady state. But such an implicit reference is often subject to criticisms in terms of environmental impact due to a lack of scientific arguments (Oenema et al., 2005). The reference may be a threshold, e.g. critical load for soil pollutant (Skeffington, 2006), a norm, e.g. water quality guidelines for nitrate, pesticides in the E.U., or a target, expressed in an absolute or in a relative way (von Wirén-Lehr, 2001; Van Cauwenbergh et al., 2007). In many cases, the definition of a reference value is not studied by scientists and is determined by the stakeholders. To our opinion it should result from the interaction between scientists and policy makers.

The subjectivity behind the choice of an absolute value is an issue and leads some authors to use references based on relative values (e.g. average of the raw values for a sample, initial value of the indicator). Such reference may enable the user to conclude that “the indicator is showing that A is better or worse than B” which does not automatically mean that “A is good, e.g. environmental friendly”.

## 3.4. Two examples of indicators to assess nitrogen losses

The following two indicators were developed to assess the importance of nitrogen losses in agrosystems. They belong to the second group of indicators presented in Section 2. The





**Figure 2.** Overview of the nitrogen indicator IN assessing the risk of nitrogen losses, nitrate, ammoniac, nitrous oxide. The main input variables involved in the different nitrogen losses are given in the boxes. Black arrays and grey arrays represent respectively the nitrogen inputs and the evolution of nitrogen in the agrosystem, i.e. gaseous emission, plant uptake, and leaching to groundwater (SMN = soil mineral nitrogen).

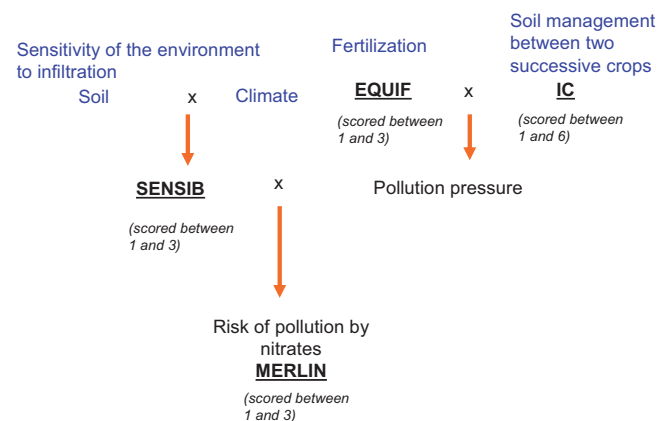
first indicator is the nitrogen indicator IN (Bockstaller and Girardin, 2001; Pervanchon et al., 2005) included in the IN-DIGO method (Bockstaller et al., 1997). It is based on a simple model simulating nitrate leaching and nitrogen gaseous emissions,  $\text{NH}_3$  and  $\text{N}_2\text{O}$ , in a quantitative way. The model outputs are transformed into scores (Fig. 2). Concept of operational model refers here to the choice of input variable based on the availability of data. Several complex inputs such as the wind speed which is a relevant variable for  $\text{NH}_3$  volatilisation are not included into the nitrogen indicator IN.

The second indicator is MERLIN (Aimon-Marié et al., 2001). This indicator can be used to assess nitrate leaching under cropping system in a qualitative way, in form of risk classes, and consists in the aggregation of three components (Fig. 3): EQUIF, an equation calculating the difference between nitrogen supply and crop requirement and assessing the risk of increase of soil mineral nitrogen at harvest due to over-fertilization, IC assessing the risk due to the management between two crops, based on the nitrogen uptake by crops and catch crops, and SENSIB assessing the leaching sensitivity of the field that results from climatic and soil conditions. SENSIB and IC components are based on contingency tables.

## 4. EVALUATION OF AN INDICATOR

### 4.1. Sensitivity analysis

A sensitivity analysis presents two major interests. First it aims at testing whether the indicator outputs are sensitive to the input variables which are known to have a strong effect,



**Figure 3.** Overview of the indicator MERLIN, combining three subindicators (EQUIF, IC and SENSIB) assessing the risk of nitrate leaching.

or whether these outputs are different for actions, e.g. cropping systems, which were found to produce different results in past studies. An example is the amount of active ingredient currently used to assess the risk of pesticide use (Levitan, 2000). This indicator is insensitive to the pesticide properties and does not differentiate two active ingredients applied at the same rate. It can not be used to assess the choice of pesticides made by farmers.

Second, sensitivity analysis allows one to analyse the effect of several input variables on the outputs of a given indicator. The results of such an analysis can be used to identify the inputs with a strong effect and those with a minor effect. The users could then decide to invest more effort on the inputs

showing a strong effect on the indicator outputs. An example can be found in Pervanchon et al. (2005) for the nitrogen indicator IN for grassland.

## 4.2. Evaluation of the quality of an indicator

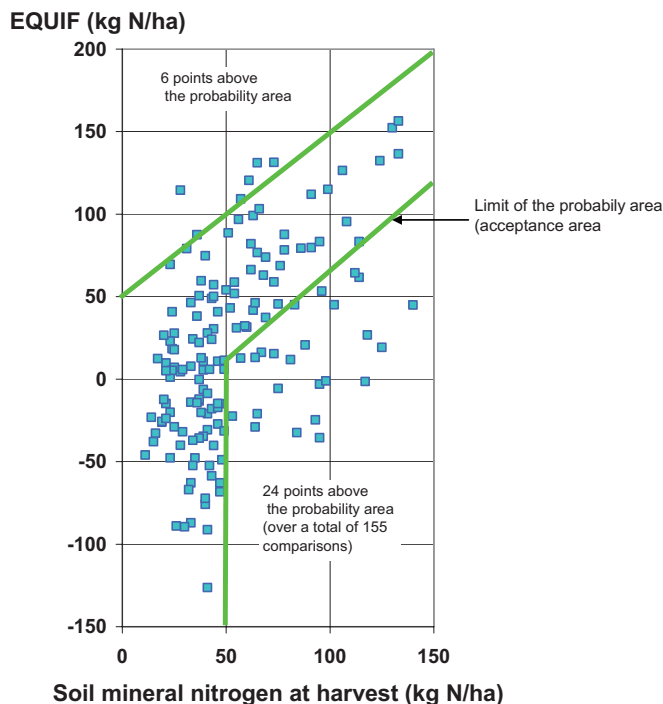
A classical approach is to evaluate the accuracy of model predictions by comparing the predicted value with observed or measured data. Some authors considered that this approach is difficult to apply to simplified indicators (Rigby et al., 2001; Reus et al., 2002). As a consequence, the accuracy of many indicators is not evaluated (Devillers et al., 2005). Bockstaller and Girardin (2003) proposed a methodological framework with three steps for the evaluation of environmental indicators which will be detailed below. This evaluation does not concern the quality of prediction only. It was recently completed by a social validation (Cloquell-Ballester et al., 2006).

### 4.2.1. Evaluation of the indicator design

This procedure consists in an evaluation of the design of an indicator by a panel of experts or by peer reviewed article. Whereas most of the publications on indicators came from the grey literature in the 1990s (Levitan, 2000), the number of scientific papers on indicators is now increasing rapidly. Such procedure allows one to check whether the design of an indicator is based on scientific knowledge and may generate outputs with a good level of accuracy. It can indicate the need of an improvement. The nitrogen indicator IN for grassland (Pervanchon et al., 2005) was strongly modified following the advices of reviewers who had rejected a first version.

### 4.2.2. Evaluation of the indicator output

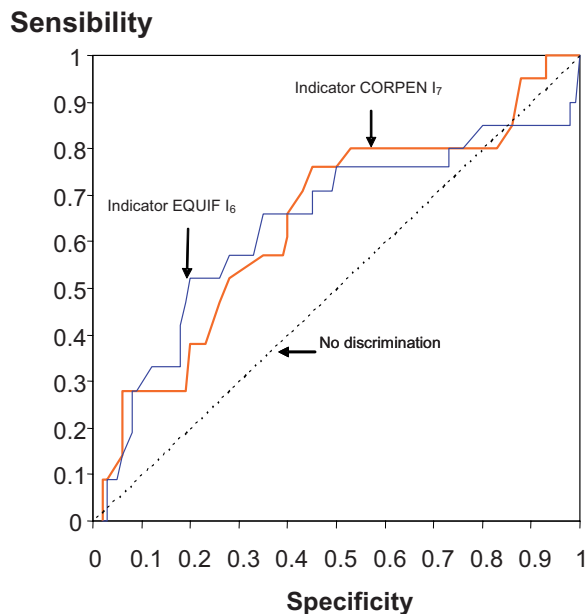
This step is based on the comparison of the indicator output with measured data. If those are not available, Bockstaller and Girardin (2003) propose alternative procedures consisting in comparison with model output or other indicator output, but this approach must be applied with caution. If the indicator is directly based on a simulation model, the model itself can be evaluated based on experimental data. For simplified indicator whose goal is not to give an accurate prediction but only some information about an environmental risk, specific approaches have been recently proposed. Bockstaller and Girardin (2003) developed a probability test which was implemented to assess the nitrogen indicator IN for grassland (Pervanchon et al., 2005) and the EQUIF sub-indicator of the indicator MERLIN (see Fig. 3). The test consists in assessing the proportion of cases in which the difference predicted value – observed value falls within a probability or acceptance area defined in function of the expected performance of the indicator and the precision of the measurements. In the example shown in Figure 4, the authors considered that the risk of leaching is low when the EQUIF output is lower than 50 kg N/ha and the soil mineral nitrogen at harvest does not exceed 50 kg/ha. Above



**Figure 4.** Probability test of the EQUIF component of the MERLIN indicator (see Fig. 3). Each point corresponds to an agricultural plot where the value of EQUIF was compared to a measurement of soil mineral nitrogen at harvest. The test consists in assessing the number of points being within a probability or acceptance area defined in function of the expected performance of the indicator and the precision of the measurements.

those thresholds, the risk increases as shown by the shape of the probability area. The test showed that a proportion of 80% of the experimental plots are in the area (Rousseau, 2003).

Makowski et al. (2005) and Primot et al. (2006) used the Receiver Operating Characteristic (ROC) methodology (Swets, 1988) to assess the ability of an indicator to discriminate between plots with high environmental risks and plots with low environmental risks. This approach was adapted to measure the accuracy agri-environmental indicators with experimental data. It consists in estimating two criteria named sensitivity and specificity. The sensitivity measures the proportion of agricultural plots with high risk correctly predicted by the indicator, the specificity measures the proportion of agricultural plots with low risk correctly predicted by the indicator. Levels of risk are determined from a measured gold standard variable and from an injury threshold. The plot of Sensitivity versus  $(1 - \text{Specificity})$  is called a “ROC curve” (see Fig. 5). The area under the ROC curve, named “AUC” for “area under curve”, is equal to the probability that the indicator values for a randomly selected pairs of agricultural plots with high and low risks will be correctly ordered. The area AUC is within the range 0–1. For a given injury threshold, perfect indicators are characterised by an AUC value equal to 1 whereas the AUC value of an useless indicator is equal to or lower than 0.5. The ROC curve can also be used to determine a decision threshold

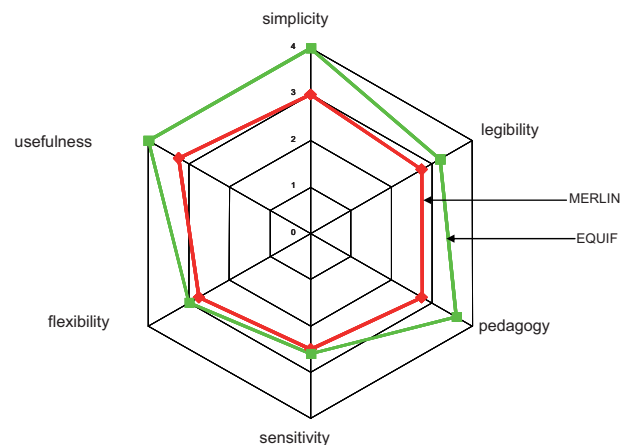


**Figure 5.** Receiver operating characteristic (ROC) curves obtained for two indicators, the EQUIF  $I_6$  (see Sect. 3.4 and Fig. 3) and the surface nitrogen balance of the CORPEN  $I_7$  and for a threshold  $Y_t = 50 \text{ kg} \cdot \text{ha}^{-1}$  of soil mineral nitrogen at harvest. The dotted line shows an area under the curve (AUC) equal to 0.5 which characterised a non discriminating indicator. Accurate indicators are characterised by high AUC values.

leading to a good compromise of sensitivity and specificity. Table I shows the AUC values estimated from 89 experimental plots located in the basin of Bruyère in France for seven nitrogen indicators. The values of seven indicators were computed and the mineral nitrogen at harvest was measured in each plot. Mineral nitrogen at harvest was considered as the gold standard and three injury thresholds were used successively. Figure 5 shows two examples of ROC curves. The results of the ROC analysis show that the EQUIF indicator ( $I_6$ ) and the Surface nitrogen balance CORPEN ( $I_7$ ) are slightly more accurate than those based on an amount of applied fertilizer ( $I_1$  to  $I_5$ ).

#### 4.2.3. Evaluation by end-users

This is the last step of the framework presented by Bockstaller and Girardin (2003). The purpose is to see whether a given indicator is used and how it is used. This step is important to identify the situations where an indicator is non applicable. It also stimulates exchanges between the developer of the indicator and its potential users. At this step, the developer can collect feed-backs from users, i.e. suggestions for improvement, problem in implementation, misunderstanding, etc. This step was implemented in the area of evaluation research to analyse how indicators are used by policy makers (Gudmundsson, 2003). Few applications were also published in cropping and farming system assessment. For example, Douguet et al. (1999) collected and analysed the reactions of farmers to indicators developed by Bockstaller et al. (1997).



**Figure 6.** Satisfaction level of users about MERLIN and EQUIF for several criteria on a scale between 1 (not satisfied) and 4 (very satisfied). Each value is an average of seven responses.

For the indicator MERLIN (see Sect. 3.4.), several quantitative criteria (Fig. 6) were quantified and qualitative feed-back recorded to seven users. The results showed a good level of satisfaction for the indicators MERLIN and for EQUIF.

## 5. DISCUSSION

Different types of indicators were presented in this article. We showed that a great diversity of indicators is available and we discussed their advantages and limitations in details. The typology presented in Figure 1 is based on the cause effect chain which was also used for the framework Pressure/State/Response of the Organisation for Economic Cooperation and Development (OECD) and for its improved version the Driving-force/Pressure/State/Impact/Response (DPSIR) of the Environmental European Agency (EEA, 2005). Payraudeau and van der Werf (2005), and Braband et al. (2003) qualified respectively the first group of indicators addressing only farmers practices as “means-based” indicator and “action-oriented” indicators whereas the other kinds on the cause-effect chain are “effect-based” indicator or “result-oriented” indicators (Fig. 1). The first group of indicators which are classified as pressure indicator failed to provide in many situations a clear link between pressure and state as it was advocated by Crabtree and Brouwer (1999). Thus, “effect-based” or “results-oriented indicators” are preferred by the authors. Of course, users have always forced to find a compromise between scientific soundness and feasibility constraints (Girardin et al., 1999), especially at the national level (Crabtree and Brouwer, 1999; Yli-Viikari et al., 2007). The cost of implementation is an important issue for many users (Romstad, 1999).

In the second part of the article, several methodological issues were reviewed; preliminary choices before the indicator development, indicator design, and indicator evaluation. All those steps imply choices and assumptions which cannot always be justified from quantitative data, but should always be



**Table I.** Values of area under curve (AUC) for the seven indicators and the three different injury thresholds of soil mineral nitrogen at harvest  $Y_t$ . The AUC under the receiver operating characteristic (ROC) curve is equal to the probability that the indicator values for a randomly selected pairs of agricultural plots with high and low risks will be correctly ordered. The value of AUC shows the ability of an indicator to discriminate between two contrasted situations. Useless indicators are characterised by an AUC value equal to or lower than 0.5. The AUC values of accurate indicators are close to 1.

Indicator	Area under the curve (AUC) estimated for the indicator for given injury thresholds $Y_t$ (kg N/ha)		
	30	40	50
$I_1$ = amount of applied nitrogen	0.58	0.57	0.55
$I_2$ = applied nitrogen + soil mineral nitrogen at winter	0.59	0.58	0.59
$I_3$ = applied nitrogen – recommended fertilizer dose	0.58	0.54	0.56
$I_4$ = applied nitrogen / grain yield	0.62	0.62	0.62
$I_5$ = [applied nitrogen + soil mineral nitrogen at winter] / grain yield	0.65	0.66	0.64
$I_6$ = soil nitrogen + apparent recovery $\times$ applied nitrogen – nitrogen requirement $\times$ grain yield (EQUIF <sup>1</sup> )	0.64	0.64	0.64
$I_7$ = applied nitrogen – nitrogen content in grain $\times$ grain yield (Surface nitrogen balance CORPEN <sup>2</sup> )	0.64	0.63	0.65

<sup>1</sup> Sub-indicator of the MERLIN indicator (see Sect. 3.4 and Fig. 3).

<sup>2</sup> Department of the French ministry of Ecology.

transparent (von Wirén-Lehr, 2001). Interactions between scientists and stakeholders should play an important role during the whole process. Regarding the design of indicators and their evaluation, the article highlights some significant progresses. The possibility of using models to design or to derive indicators goes beyond the opposition between model and indicator discussed by Bockstaller et al. (1997). For the evaluation of indicators by comparison with measured data, two methods have been proposed which can be complementary. A probability test can be used to provide information about the relation between indicators and measurements. The ROC method can be used to assess the ability of an indicator to discriminate between situations with high and low environmental risks. It can also be used to define decision thresholds from experimental data in function of sensitivity and specificity target values.

Some issues were not addressed in this article but will deserve more attention in the next few years. The choice of the scale for calculating indicator outputs is an important issue and should be discussed in relation with the type of impact, the status of the indicator on the cause-effect chain. For example, water quality indicators should be used at the scale of the water catchments or for a landscape. Emissions can be assessed at lower scale of the cropping and farming systems. For indicator assessing emissions, results can be upscaled by aggregation of results obtained by calculation of an average value at higher scale weighted by the size or the number of entity at lower scale. Such aggregation at higher scales like a nation is not relevant for local impact, e.g. water quality, erosion, whereas it is possible for global impact, e.g. greenhouse gases. Upscaling requires some statistical skills for data management but must also integrate new processes (Stein et al., 2001) and new environmental components (e.g non cropped area).

A second issue concerns the evaluation of the indicators, and more precisely the uncertainty linked to the indicator. Authors working on Life Cycle Analysis approaches are con-

cerned by this issue (Basset-Mens et al., 2006). Some addressed it using fuzzy logic approach (Ardente et al., 2004). Uncertainty was analysed for nitrogen balances by using fuzzy logic (Mertens and Huwe, 2002) or Monte-Carlo approaches (Oenema et al., 2003). Finally, a third issue concerns the interpretation and use of the indicator outputs (Yli-Viikari et al., 2007). Recommendations about the significance of the scores, about the uncertainty of the results, the relevance of the reference level should be given to the users to help them to interpret the results.

To facilitate the interpretation of set of results, aggregation is often used. This is true for the two types of aggregation regarding upscaling procedures (see above) and composite or multi-criteria aggregation where indicator related to different themes are combined. The relevance of aggregation is often discussed because of the loss of information but also due to the methodological problems it raises. A major problem is “adding apple and pear” in the case of composite indicators which can appear in scoring method (Rigby et al., 2001). Several method are available to avoid this problem like the normalization technique in monetary unit or physic unit, the multivariate approach (Nardo et al., 2005), or the decision trees using fuzzy logic (van der Werf and Zimmer, 1998; Phillis and Andriantiatsaholainiaina, 2001). An alternative consists in using the multi-criteria methods based on an outranking procedure (Arondel and Girardin, 2000; Hayashi, 1998). One shortcoming of this last family of methods is the principle of outranking based on relative comparisons and not on an absolute assessment. The use of weightings procedure in aggregation as well as in multi-criteria methods is also often criticized due to its subjectivity. This cannot be totally avoided but should be transparent. Andreoli et al. (1999) proposed guidelines for this problem. Sensitivity analysis may be useful to assess the effects of weighting. Such analysis will also help to cope with compensation and trade-off between sub-indicators

in a composite indicator (Nardo et al., 2005). We advise to use both aggregated and individual indicators.

Several issues mentioned in this article require interactions with users, from the elaboration and selection of indicators, to the interpretation of results. The users can be involved at the beginning of the elaboration of an indicator or a set of indicators in a procedure of participative research. In this case, the characteristics and required qualities of indicators are defined together by end-users and scientists. At the end of the development, an evaluation by end-users may bring new information from end-users to scientist to improve the indicator. With the growing number of indicators and methods available for the end-users, the question will shift from "how to elaborate an indicator?" to "which indicators?", so that they need comparative information as in Reus et al. (2002), Devillers et al. (2005), or methodological help to compare indicators. More methodological research is needed on this issue of indicator comparison.

## 6. CONCLUSION

Many indicators are available to help agronomists and stakeholders working on the assessment of sustainability of farming and cropping systems. This article presents a typology of environmental indicators and discusses their advantages and limitations. In many cases, only few data are available, especially at regional or higher levels. Only simple indicators based on farmer's practices can be used in such cases. These indicators generally present a low quality of prediction. They can be combined in order to improve their accuracy, but the use of multiple indicators is often complicate in practice. Efficient methods for integrating various processes are still needed. When input data on soil and farmers' practices are available, indicators based on operational model like those presented in this article for nitrogen losses can be useful to analyse the effects of various factors related to soil, climate, and cropping systems. Such indicators are still lacking in several areas, notably to assess the impact of agriculture on biodiversity.

In cropping system experiments, measured indicators and model-based indicators should be both used. Model-based indicators are often required in this context because all the variables of interest cannot always be measured like, for example, gaseous emissions or pesticide losses. In any case, we advised agronomists and environmentalists to use the methodological framework proposed in this paper to design indicators. Issues like scales and upscaling procedures, uncertainty analysis, interpretation of the results, interaction with indicator users and comparison of indicators should be the object of more research work during the next few years.

## REFERENCES

- Aimon-Marié F., Angevin F., Guichard L. (2001) Une méthode agronomique pour apprécier les risques de pollution diffuse par les nitrates d'origine agricole, Chambre d'Agriculture de Charente Maritime, La Rochelle, p. 29.
- Andreoli M., Rossi R., Tellarini V. (1999) Farm sustainability assessment: some procedural issues, *Landscape Urban Plan.* 46, 41–50.
- Ardente F., Beccali M., Cellura M. (2004) FALCADE: a fuzzy software for the energy and environmental balances of products, *Ecol. Model.* 176, 359–379.
- Aronel C., Girardin P. (2000) Sorting cropping systems on the basis of their impact on groundwater quality, *Eur. J. Oper. Res.* 127, 476–482.
- 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.
- Basset-Mens C., van der Werf H.M.G., Durand P., Leterme P. (2006) Implications of uncertainty and variability in the life cycle assessment of pig production systems, *Int. J. LCA* 11, 298–304.
- Bockstaller C., Girardin P. (2001) "IN", an indicator to assess nitrogen losses in cropping systems, 11th Nitrogen Workshop, INRA, Reims, pp. 419–420.
- 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.
- Bossel H. (1999) Indicators for sustainable development: Theory, method, applications. IISD International Institute of Sustainable Development, Winnipeg, Manitoba USA, p. 125.
- Braband D., Geier U., Kopke U. (2003) Bio-resource evaluation within agri-environmental assessment tools in different European countries, *Agr. Ecosyst. Environ.* 98, 423–434.
- Brentup F., Kusters J., Kuhlmann H., Lammel J. (2004) Environmental impact assessment of agricultural production systems using the life cycle assessment methodology - I. Theoretical concept of a LCA method tailored to crop production, *Eur. J. Agron.* 20, 247–264.
- Brown C.D., Hart A., Lewis K.A., Dubus I.G. (2003) P-EMA (1): simulating the environmental fate of pesticides for a farm-level risk assessment system, *Agronomie* 23, 67–74.
- Carignan V., Villard M.A. (2002) Selecting indicator species to monitor ecological integrity: A review, *Environ. Monit. Assess.* 78, 45–61.
- Clergué B., Amiaud B., Pervanchon F., Lasserre-Joulin F., Plantureux S. (2005) Biodiversity: function and assessment in agricultural areas. A review, *Agron. Sustain. Dev.* 25, 1–15.
- Cloquell-Ballester V.A., Cloquell-Ballester V.A., Monterde-Díaz 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.
- Crabtree J.R., Brouwer F.M. (1999) Discussion and conclusions, in: Brouwer F.M., Crabtree J.R. (Eds.), *Environmental indicators and agricultural policy*, CAB International, Wallingford, pp. 279–285.
- Delbaere B., Serradilla A.N.E. (2004) Environmental risks from agriculture in Europe: Locating environmental risk zones in Europe using agri-environmental indicators, EC NC-European Centre for Nature Conservation, Tilburg, The Netherlands, p. 184.
- 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.
- Douguet J.M., O'Connor M., Girardin P. (1999) Validation socio-économique des Indicateurs Agro-écologiques. C3ED, Université de Versailles Saint-Quentin-en-Yvelines, Guyancourt, France, p. 73.
- Durand P., Gascuel-Odoux C., Cordier M.O. (2002) Parametrisation of hydrological models: a review and lesson learned from studies on an agricultural catchment (Naizin, France), *Agronomie* 22, 217–228.

- 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 (Denmark), p. 128.
- Enea M., Salemi G. (2001) Fuzzy approach to the environmental impact evaluation, *Ecol. Model.* 136, 131–147.
- Flemström K., Carlson R., Erixon M. (2004) Relationships between Life Cycle Assessment and Risk Assessment – Potentials and Obstacles, Chalmers University of Technology, Göteborg (Sweden), p. 82.
- Freyer B., Reisner Y., Zuberhuhler D. (2000) Potential impact model to assess agricultural pressure to landscape ecological functions, *Ecol. Model.* 130, 121–129.
- Garcet J.D.M., Ordóñez A., Roosen J., Vanclooster M. (2006) Metamodelling: Theory, concepts and application to nitrate leaching modelling, *Ecol. Model.* 193, 629–644.
- Girardin P., Bockstaller C., van der Werf H.M.G. (1999) Indicators: tools to evaluate the environmental impacts of farming systems, *J. Sustain. Agr.* 13, 5–21.
- 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.
- Gras R., Benoit M., Deffontaines J.P., Duru M., Lafarge M., Langlet A., Osty P.L. (1989) *Le fait technique en agronomie, Activité agricole, concepts et méthodes d'étude*, Institut National de la Recherche Agronomique, L'Hamartan, Paris, France.
- Gudmundsson H. (2003) The policy use of environmental indicators – learning from evaluation research, *J. Transdisciplinary Environ. Studies* 2, 1–11.
- Hansen J.W. (1996) Is agricultural sustainability a useful concept? *Agr. Syst.* 50, 117–143.
- Hayashi K. (1998) Multicriteria aid for agricultural decisions using preference relations: Methodology and application, *Agr. Syst.* 58, 483–503.
- 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.
- Hess B., Schmid H., Lehmann B. (1999) Umweltindikatoren – Scharnier zwischen Ökonomie und Ökologie, *Agrarforsch.* 6, 29–32.
- Hülsbergen K.J. (2003) Entwicklung und Anwendung eines Bilanzierungsmodells zur Bewertung der Nachhaltigkeit landwirtschaftlicher Systeme, Shaker Verlag (Halle, Univ., Habil.-Schr., 2002), Aachen.
- Jollands N. (2006) How to aggregate sustainable development indicators: a proposed framework and its application, *Int. J. Agric. Resources Governance Ecol.* 5, 18–34.
- Levitan L. (2000) "How to" and "why": assessing the enviro-social impacts of pesticides, *Crop Prot.* 19, 629–636.
- 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.
- Makowski D., Taverne M., Bolomier J., Ducarne M. (2005) Comparison of risk indicators for sclerotinia control in oilseed rape, *Crop Prot.* 24, 527–531.
- 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.
- Maurizi B., Verrel J.-L. (2002) Des indicateurs pour des actions de maîtrise des pollutions d'origine agricole, *Ingénieries* 30, 3–14.
- Merkle A., Kaupenjohann M. (2000) Derivation of ecosystemic effect indicators - method, *Ecol. Model.* 130, 39–46.
- Mertens M., Huwe B.U. (2002) FuN-Balance: a fuzzy balance approach for the calculation of nitrate leaching with incorporation of data imprecision, *Geoderma* 109, 269–287.
- Meyer-Aurich A. (2005) Economic and environmental analysis of sustainable farming practices – a Bavarian case study, *Agr. Syst.* 86, 190–206.
- Mitchell G., May A., McDonald A. (1995) PICABUE: a methodological framework for the development of indicators of sustainable development, *Int. J. Sustain. Dev. World Ecol.* 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.
- Oenema O., Kros H., de Vries W. (2003) Approaches and uncertainties in nutrient budgets: implications for nutrient management and environmental policies, *Eur. J. Agron.* 20, 3–16.
- 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.
- Pervanchon F., Bockstaller C., Bernard P.Y., Peigné J., Amiaud B., Vertès F., Fiorelli J.L., Plantureux S. (2005) A novel indicator of environmental risks due to nitrogen management on grasslands, *Agr. Ecosyst. Environ.* 105, 1–16.
- Phillis Y.A., Andriantiatsaholainaina L.A. (2001) Sustainability: an ill-defined concept and its assessment using fuzzy logic, *Ecol. Econ.* 37, 435–456.
- Prato T. (2005) A fuzzy logic approach for evaluating ecosystem sustainability, *Ecol. Model.* 187, 361–368.
- Primot S., Valantin-Morison M., Makowski D. (2006) Predicting the risk of weed infestation in winter oilseed rape crops, *Weed Res.* 46, 22–33.
- Purtauf T., Thies C., Ekschmitt K., Wolters V., Dauber J. (2005) Scaling properties of multivariate landscape structure, *Ecol. Indic.* 5, 295–304.
- Reus J., Leenderste P., Bockstaller C., Fomsgaard I., Gutsche V., Lewis K., Nilsson C., Pussemier L., Trevisan M., van der Werf H., Alfaro 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.
- Rigby D., Woodhouse P., Young T., Burton M. (2001) Constructing a farm level indicator of sustainable agriculture practice, *Ecol. Econ.* 39, 463–478.
- 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.
- Riley J. (2001c) Multidisciplinary indicators of impact and change – Key issues for identification and summary, *Agr. Ecosyst. Environ.* 87, 245–259.
- Romstad E. (1999) Theoretical considerations in the development of environmental indicators, in: Brouwer F.M., Crabtree J.R. (Eds.), *Environmental indicators and agricultural policy*, Wallingford, CAB International, 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, September 4–6, 2006, pp. 707–708.
- Rousseau M.L. (2003) L'évaluation de MERLIN, une méthode utilisée en Poitou-Charentes constituée d'indicateurs agro-environnementaux (EQUIF, IC, SENSIB), ESA Angers, p. 75.

- Sadok W., Angevin F., Bergez J.-E., Bockstaller C., Colomb B., Guichard L., Reau R., Doré T. (2007) Ex ante assessment of the sustainability of alternative cropping systems: guidelines for identifying relevant multi-criteria decision aid methods, *Agron. Sustain. Dev.* DOI: 10.1051/agro: 2007043.
- Shan Y., Paull D., McKay R.I. (2006) Machine learning of poorly predictable ecological data, *Ecol. Model.* 195, 129–138.
- Silvert W. (2000) Fuzzy indices of environmental conditions, *Ecol. Model.* 130, 111–119.
- Skeffington R. (2006) Quantifying uncertainty in critical loads: (A) literature review, *Water Air Soil Pollut.* 169, 3–24.
- Stein A., Riley J., Halberg N. (2001) Issues of scale for environmental indicators, *Agr. Ecosyst. Environ.* 87, 215–232.
- Swets J. (1988) Measuring the accuracy of diagnostic systems, *Science* 240, 1285–1293.
- ten Berge H.F.M., Burgers S.L.G.E., Schröder E.J., Hofstad E.J. (2002) 'Partial balance'-regression models for  $N_{min,H}$ , in: ten Berge H.F.M. (Ed.), A review of potential indicators for nitrate loss from cropping and farming systems in the Netherlands, *Plant Research International B.V., Wageningen (The Netherlands)*, pp. 25–60.
- Van Cauwenbergh N., Biala K., Biolders C., Brouckaert V., Franchois L., Garcia Ciudad V., Hermy M., Mathijs E., Muys B., Reijnders J., Sauvenier X., Valckx J., Vanclooster M., Van der Veken B., Wauters E., Peeters A. (2007) SAFE-A hierarchical framework for assessing the sustainability of agricultural systems, *Agr. Ecosyst. Environ.* 120, 229–242.
- van der Werf H.G.M., Petit J. (2002) Evaluation of environmental impact of agroculture 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.
- von Wirén-Lehr S. (2001) Sustainability in agriculture – an evaluation of principal goal-oriented concepts to close the gap between theory and practice, *Agr. Ecosyst. Environ.* 84, 115–129.
- Yli-Viikari A., Hietala-Koivu R., Huusela-Veistola E., Hyvonen T., Perala P., Turtola E. (2007) Evaluating agri-environmental indicators (AEIs) – Use and limitations of international indicators at national level, *Ecol. Indic.* 7, 150–163.