Benchmarking sustainability farm performance at different levels and for different purposes: elucidating the state of the art

Coteur, I.[1], Marchand, F.[1,2], Van Passel, S.[2], Schader, C.[3], Debruyne, L.[1], Wustenberghs, H.[1] and Keszthelyi, S.[4]

[1] Institute for Agricultural and Fisheries Research (ILVO)-Social sciences Unit, Burg. Van Gansberghelaan 115, box 2; 9820 Merelbeke, Belgium

[2] University of Antwerp, Ecosystem Management Research Group and IMDO, Universiteitsplein 1, 2610 Wilrijk, Belgium

[3] FiBL - Research Institute of Organic Agriculture, Frick, Switzerland

[4] Research Institute of Agricultural Economics, Budapest

Keywords: integrated sustainability assessment, benchmarking, tool characteristics

Abstract

Multiple indicators for agroecosystems, sustainable land management, social development, rural livelihoods, biodiversity, etc. were already developed many years ago (Riley, 2001). Nowadays many of those indicators are used in a more holistic way, encompassing several or all of the aspects mentioned. However, this abundancy of frameworks, tools and metrics for agricultural sustainability assessment is still growing (Pope *et al*., 2013; Schindler, 2015). How does one navigate between benchmarking systems and sustainability assessment tools? What are the key characteristics to describe frameworks, metrics and tools that may facilitate the choice between them? How can one select the most appropriate for one's purpose?

Our objective is to provide starting points to answer these questions. We performed a literature review regarding the characteristics proposed to discern metrics and tools. We used state of the art results from the OECD TempAg network, who inventoried integrated sustainability assessment tools and metrics designed for different purposes, to divide existing metrics and tools according to these characteristics focusing on the purpose, level and end-user. This paper first addresses conceptual aspects regarding sustainability assessment. Second, it describes the method used to define the characteristics, the characteristics themselves and third it shows the division of the tools. Our research resulted in a list of 25 characteristics, which were grouped into general assessment related information, information related to stakeholder participation and indicators related information. The division of tools and metrics according to these characteristics raises new questions and starting points for future research and helps us to refine our research questions.

1. Introduction

Increasing attention toward sustaining the environment in the early '90s has led to the development of tools and metrics to assess sustainable development (Riley, 2001). These tools and metrics ranged from indicator lists, assessment models and indexes (Binder et al., 2010). They were developed for one or more specific themes or issues, had different aims and were related to different systems (Bockstaller & Guichard, 2009; Riley, 2001). At first, the focus laid on environmental aspects (Rigby and Caceres, 1997 in Binder et al., 2010; Pope et al., 2004), but over time these tools were used in more holistic and integrated frameworks (Binder et al., 2010). Sustainability assessment has become an important aid in the process toward sustainability (Pope et al., 2004). It is defined by many authors and can be seen as a process which directs decision-making towards sustainability, integrates sustainability concepts into decision making or operationalizes sustainable development as a guide for decision making by identifying the future consequences of current and planned actions (Bond et al., 2012; Hugé et al., 2013; Pope, 2006). Sustainability assessment tools and frameworks are developed to assess sustainability and facilitate sustainability assessment (Coteur et al., *in press*, derived from Gasparatos and Scolobig, 2012 and Ness et al., 2007). However, literature on sustainability assessment and sustainability assessment tools to support decision making is ever-expanding (Binder et al., 2010; Bockstaller and Guichard, 2009; Carof et al., 2013; Gasparatos and Scolobig, 2012; Marchand et al., 2014; Ness et al., 2007; Schindler et al., 2015). Many divers processes are described as sustainability assessment due to its broad definition (Pope et al., 2004).

Questions arise on how to navigate between these tools, what are their key characteristics and how can one select the most appropriate for one's purpose? However, literature is lacking regarding tool choice and effective use of tools and methodologies (De Ridder et al., 2007; Gasparatos & Scolobig, 2012). The objective of this research is to contribute to this gap in literature by providing a starting point to answer these questions. We performed a literature review regarding the characteristics proposed to discern metrics and tools. Furthermore, we used state of the art results from the OECD TempAg network, who inventoried integrated sustainability assessment tools and metrics designed for different purposes, to divide existing metrics and tools according to these characteristics focusing on the purpose, level and end-user. This paper first describes key issues regarding sustainability assessment, it continues by describing the characteristics of assessment tools and metrics and ends with a discussion on existing assessment tools and metrics analysed according to the described characteristics.

2. Key issues regarding sustainability assessment

Many authors have already discussed key-issues regarding the design and use of a sustainability assessment (e.g. Binder et al., 2010; Gasparatos & Scolobig, 2012; Gibson, 2006; Ness et al., 2007; Pope et al., 2004; Weaver & Rotmans, 2006). We will highlight some of the aspects of sustainability assessment, but like to stress that this list is not all-inclusive. One of the key-issues is the contested meaning of sustainability and sustainable development (Bond et al., 2013; Hopwood et al., 2005; Pope et al., 2004; Waas et al, 2011). Bond and Morrison-Saunders (2013) (in Bond et al., 2013) describe five critical debates, issues related to sustainability assessment, two of them related to the concept of sustainability. They state that the meaning of sustainability should be formulated for every assessment, taking into account the context in which it occurs (Bond & Morrison-Saunders, 2013). The other critical debates are related to the indicator selection (holism versus reductionism), the contested time horizons and spatial boundaries and the design of the assessment process itself as well as its outcomes (process versus outcome). Also Binder et al. (2010) highlight the need for a well-defined normative dimension of sustainability assessments, including the concept of sustainability(C R Binder et al., 2010). However, pluralism, a wide variety of views, should be seen as an opportunity and is an essential aspect of sustainability assessment (Pope & Morrison-Saunders, 2013). Therefore, Bond and Morrison-Saunders (2013) conclude, among other things, that communication to stakeholders about these debates or issues is crucial to create certainty and improve the credibility of the sustainability assessment.

Not only the assessment itself, but also the tools and frameworks used to facilitate these assessments are subject to variety. These assessment tools can have different purposes such as certification, communication (non-committal), reporting to policy makers (obligatory), firm development or research. Many assessment tools are designed to assess at a specific level of scale. Some will assess the farm level or field level, others the sector level, production system level, regional level or the land unit scale. Literature shows that these different levels also suggest different end-users (Van Passel & Meul, 2010). Tools assessing the farm level will be mostly used by farmers as results can be used to improve the sustainable performance of the individual farm. On the other hand, tools assessing sector and regional level are most interesting for policy makers as policy measures are drawn up at these levels. Other end-users of tools can be extension workers, researchers, NGOs or actors of the supply chain. Assessment tools can also focus on different aspects of sustainability such as economic, social, cultural, environmental or governance aspects. This list of differences is not all-inclusive as many differences between tools occur. This variety of characteristics should however be taken into account when developing or selecting sustainability assessment tools and during the design of a sustainability assessment.

Furthermore, assessment tools are made up of indicators or metrics. Indicators are used to describe and determine the state or presence of a complex system (Steunpunt Duurzame Landbouw, 2006;

UNAIDS, 2010). They measure performance or reflect changes related to activities, projects or programmes (UNAIDS, 2010), without necessarily measuring the state of the system itself (Steunpunt Duurzame landbouw, 2006). These indicators can be quantitative or qualitative and their results can be visually or numerically integrated. Visual integration means that the indicator results are presented together within a table or diagram. Numerical integration combines the indicator results to present it as a single index or composite indicator (Gómez-Limón & Riesgo, 2009; Van Passel et al., 2007). Data source, the way of integration, weighing of the indicators and other factors are important aspects of integration and transparency is needed when dealing with composite indicators (Van Passel et al., 2007). A system can be represented in a holistic way by using many divers indicators or in a reductionist way by using just a few indicators to assess a whole system.

When we want to gauge or compare the performance of a system, assessed with an assessment tool, we can use benchmarks. Benchmarking means comparing your own performance against a standard or with the performance of others. It involves continuous measuring, analysing and taking action to improve our performance (Poppe & van Asseldonk, 2015). There are different ways of setting a benchmark such as a predefined value from literature, regulatory standards or a benchmark based on the performance of other systems (Binder et al., 2010; Poppe and van Asseldonk, 2015).

3. Characteristics for assessment system description

As the previous section shows, the variety among tools is immense and there are numerous ways to categorize frameworks, metrics and tools for agricultural sustainability assessment. However, the question remains how to navigate and choose between these tools (de Ridder et al., 2007; Gasparatos and Scolobig, 2012). What are the key characteristics that may facilitate the choice between these tools?

In the context of the TempAg research collaboration on sustainable temperate agriculture an in-depth literature review was performed regarding the characteristics to discern metrics and tools. This specific research collaboration focusses on three themes and the literature review fits within the first theme "Delivering Resilient Agricultural Production Systems at Multiple Spatial and Temporal levels". A first research question posed within this theme is "How can sustainability frameworks, metrics and tools and their implementation be enhanced to futureproof agricultural decision making at multiple levels on multiple scales?". To answer this question an inventory of existing frameworks and tools was developed and each tool was analysed on the basis of the list of characteristics. The frameworks, metrics and tools that were selected are specific to agriculture or applicable to agriculture, developed in and/or applicable in temperate climates and designed to assess sustainability in an integrated way (at least three dimensions – economic, environmental and social). Emphasis was somewhat put on farm level assessments.

The literature review resulted in a list of 70 characteristics. As the meaning and denomination of certain characteristics can vary between authors, characteristics with high similarity were clustered and working definitions were formulated. The list has been reduced to 25 essential characteristics, for which definitions were univocally formulated. These 25 characteristics, presented in table 2, were grouped according to general assessment related information, information related to stakeholder participation and indicators related information.

Table 1:Characteristics for assessment system description

ASSESSMENT RELATED CHARACTERISTICS	
Characteristic	Definition
origin	developed in which country or countries
initiative	developed on the initiative of ?
dating	year of development
scope of assessment	dimensions of sustainability considered (economic, environmental, social, governance, cultural)
perspective on sustainability	perspective on sustainability within scope (definition of sustainability used)
primary purpose of the assessment	the intended function of the tool: reporting (obligatory), communication (non-committal), firm development, research, certification,…
level of assessment	Spatial scale of the assessment: field, farm, industry, chain, national/regional, landscape, global, product,…
sector scope	The assessed farm type or production type: general (applicable to all agricultural/food products or farm types; applicable to specific products or farm types (+ define which one)
system representation	Is the system represented in a reductionist (few indicators are used to assess the sustainability of a whole system) or holistic (reflects the complexity of a system by using many divers indicators) way?
applying user	The one applying the assessment: individual farmers, extension workers, policy makers, researchers,… or a combination: farmer and extension (Schindler *et al*., 2015)
end-user of results	The end-user of the results: individual farmer, farmers in discussion groups, extension workers, policy makers, researchers,… or a combination: farmer + extension/farmers in discussion groups (Bockstaller *et al*., 2015; Schindler *et al*., 2015)
method of data collection	method of data collection: interview (individual farmer + extension worker); audit (control system); self-assessment (tools that can be used and interpreted individually); other
aggregation & weighting	Are the indicator scores aggregated? Yes, No; If yes, is it a weighted aggregation? To which level?; If yes to weighing, method of weighing?
transparency	Are there reports/documents available for users regarding: content, purpose, method of assessment, indicator scores, interpretation of results, other?
level of implementation	Is the assessment being used, implemented? If yes; specify: only on a project basis, commercially used, used by farmers, used for certification, other

Table 1 (continued)

STAKEHOLDER PARTICIPATION	
Characteristic	Definition
What was the type of stakeholder participation for every phase of the assessment?	
stakeholder participation when?	Following the 6 stages defined by Binder *et al*. (2010): *(1) Preparatory phase:* defining context, goal and challenges; *(2)* Indicator *selection:* choosing the appropriate sustainability indicators, taking decisions on including interactions between indicators and how to weight indicators; *(3) Indicator measurement*: quantification of indicators and processes (use of statistical data, surveys or categorized qualitative data); *(4) Aggregation of indicators:* taking decisions on whether or not to aggregate indicators, to which extent and how; *(5) Applicability* of the assessment results: the process of getting the generated knowledge ready for utilization in practice; *(6) Follow-up:* reporting results, developing management advice, monitoring of indicators over time.
stakeholder participation who?	Who was involved? (farmers, extension workers (advisors), researchers, policy makers, civil society,…)
stakeholder participation how?	What type of stakeholder participation? (interviews, focus groups, workshops, other)
time for data collection	Time requirement for data collection (categories: < 2 h; 2-4 h; 1 day; 2 days; > 2 days)
INDICATOR RELATED CHARACTERISTICS - ACCURACY OF METHOD CALCULATION	
Characteristic	Definition
indicator type	Primarily quantitative; primarily qualitative; equally quantitative and qualitative indicators
level of data input	Are the data needed to complete the assessment at field level, farm level, product level, region level or other?
data source	type of data used: accountancy, farmers' knowledge, expert information, field practices, site practices, other
number of topics	What is the number of topics for this dimension? • Number of themes • Number of indicators
reliability of data input	Are the data used for assessing correct and reliable? Yes, for all indicators within this dimension; yes, for most indicators of this dimension; no, data input for many indicators is doubtful
validation of method calculation	Are the calculation methods validated? If yes, what type of validation was used?
scoring system	What kind of scoring system was used for scoring the indicators of this dimension? benchmarks: which method is used?; expert based scoring: which method is used?; scoring from literature; other

4. A list of initiatives divided by primary purpose, end-user and level of assessment

The preliminary results of the TempAg inventory show a large variation in development and content of the tools. Table 2 shows a variation of initiatives divided by the primary purpose of the tool (reporting, firm development, communication, research and certification), the end-user (farmer and policy) and the level of assessment (firm, sector and regional level). Farmer and policy were chosen as end-user, because their aims (developing a farm or building/redirecting legislation) might be furthest apart.

As our original selection of initiatives focused somewhat on the farm level assessments, we see a larger amount of tools for the farmer as end-user. We said before that a specific level of scale suggests a different end-user. Results from Van Passel and Meul (2011) show significant differences between these levels as tools which assess at sector and regional level are only used by policy makers. This does however not show from our preliminary data results and many tools have multiple assessment levels, serving both the policy maker as well as the farmer (e.g. COSA indicators, NZSD, FAO-SAFA, SMART). These tools also have different purposes, ranging from firm development to research. These observations should be further investigated. Why do so many tools claim at serving both end-users and are these tools really used by both end-users? What does assessing at firm, sector or regional level mean for these tools? How do they fulfil these purposes and what methods do they use?

In any way, we need to gain more insight in the differences between these tools. If we do need different tools for farmers and policy makers, is it necessary or feasible to align data collection and therefore make the assessment more efficient? If there are tools of which the results can be used by both the farmer and policy maker, how are these results presented and used? Is interpretation of the results more difficult if it needs to serve both the farmer and the policy maker?

5. Conclusion

Questions arise on how to navigate between sustainability assessment tools. What are their key characteristics and how can one select the most appropriate for one's purpose? This research resulted in a list of 25 essential characteristics to discern tools and metrics. These characteristics were grouped according to general assessment related information, information related to stakeholder participation and indicators related information. It is a first starting point to guide tool selection as more insight is gained when analysing tools according to the characteristics. Furthermore, we divided a number of tools and metrics based on the purpose of the assessment, its level and the end-user. This preliminary result showed that a number of tools can be used by farmers and policy makers, used at different levels and for different purposes. However, these results pose new questions for future research. What is the difference between tools designed for a farmer or a policy maker? Do they use the same data source and how does data collection work? Are results presented in a different way and how do these end-users use the results in decision making?

Table 2: List of initiatives divided according to primary purpose, end-user and level of assessment

		FARMER	POLICY
REPORTING	firm level	COSA Indicators FAO-SAFA FLINT FtoM GRI INSPIA LEAF-SFR NZSD SAI-FSA2.0 SMART SPA	COSA Indicators FAO-SAFA FLINT GRI NZSD SMART
	sector level	COSA Indicators FAO-SAFA NZSD SMART	COSA Indicators FAO-SAFA FLINT NZSD SMART
	regional level	NZSD	FLINT NZSD
FIRM DEVELOPMENT	firm level	BJCD BRP COSA Indicators DEXiFruits DEXiPM EISA FAO-SAFA INSPIA KSNL LEAF-SFR MESMIS MOTIFS NZSD ORC-FAS RISE SAI-FSA2.0 SAN-SAS ScalA SMART Veldleeuwerik	BJCD COSA Indicators DEXiPM FAO-SAFA NZSD ScalA SMART
	sector level	COSA Indicators FAO-SAFA NZSD SMART	COSA Indicators FAO-SAFA NZSD SMART
	regional level	MESMIS NZSD	NZSD

Table 2: (continued)

		FARMER	POLICY
COMMUNICATION	**firm level**	FAO-SAFA INSPIA KSNL LEAF-SFR MOTIFS RISE SAI-FSA2.0 SMART	FAO-SAFA FLINT SMART
	sector level	FAO-SAFA SMART	FAO-SAFA SMART
	regional level	FoPIA	FoPIA
RESEARCH	**firm level**	COSA Indicators DEXiPM MESMIS SMART	COSA Indicators DEXiPM FLINT SEAMLESS SMART SVA
	sector level	COSA Indicators SMART	COSA Indicators FLINT SMART
	regional level	FoPIA MESMIS	FLINT FoPIA SEAMLESS TOA-MD 5.0 model
CERTIFICATION	**farm level**	GlobalGAP KSNL LEAF-Marque SAN-SAS	
	sector level		
	regional level		

6. References

Binder, C. R., Feola, G., & Steinberger, J. K. (2010). Considering the normative, systemic and procedural dimensions in indicator-based sustainability assessments in agriculture. *Environmental Impact Assessment Review*, *30*(2), 71–81. doi:10.1016/j.eiar.2009.06.002

Binder, C. R., Feola, G., & Steinberger, J. K. (2010). Considering the normative, systemic and procedural dimensions in indicator-based sustainability assessments in agriculture. *Environmental Impact Assessment Review*, *30*(2), 71–81. doi:10.1016/j.eiar.2009.06.002

Bockstaller, C., & Guichard, L. (2009). Comparison of methods to assess the sustainability of agricultural systems: A review. *Agron. Sustain. Dev.*, *29*, 223–235. doi:10.1051/agro

Bond, A. J., & Morrison-Saunders, A. (2013). Challenges in determining the effectiveness of sustainability assessment. In A. J. Bond, A. Morrison-Saunders, & R. Howitt (Eds.), *Sustainability assessment: pluralism, practice and progress* (pp. 37–50). Routledge.

Bond, A. J., Morrison-Saunders, A., & Howitt, R. (2013). *Sustainability assessment: pluralism, practice and progress*. (A. J. Bond, A. Morrison-Saunders, & R. Howitt, Eds.). Routledge.

Carof, M., Colomb, B., & Aveline, a. (2013). A guide for choosing the most appropriate method for multi-criteria assessment of agricultural systems according to decision-makers' expectations. *Agricultural Systems*, *115*, 51–62. doi:10.1016/j.agsy.2012.09.011

De Ridder, W., Turnpenny, J., Nilsson, M., & Von Raggamby, A. (2007). A Framework For Tool Selection And Use In Integrated Assessment For Sustainable Development. *Journal of Environmental Assessment Policy and Management JEAPM*, *9*(04), 423–441. Retrieved from http://ideas.repec.org/a/wsi/jeapmx/v09y2007i04p423-441.html

Gasparatos, A., & Scolobig, A. (2012). Choosing the most appropriate sustainability assessment tool. *Ecological Economics*, *80*, 1–7. doi:10.1016/j.ecolecon.2012.05.005

Gibson, R. B. (2006). BEYOND THE PILLARS : SUSTAINABILITY ASSESSMENT AS A FRAMEWORK FOR EFFECTIVE INTEGRATION OF SOCIAL , ECONOMIC AND ECOLOGICAL CONSIDERATIONS IN SIGNIFICANT DECISION-MAKING The core argument here is quite simple . Because sustainability is an essentially int, *8*(3), 259–280.

Gómez-Limón, J. a, & Riesgo, L. (2009). Alternative approaches to the construction of a composite indicator of agricultural sustainability: An application to irrigated agriculture in the Duero basin in Spain. *Journal of Environmental Management*, *90*(11), 3345–62. doi:10.1016/j.jenvman.2009.05.023

Hopwood, B., Mellor, M., & Brien, G. O. (2005). Sustainable Development - Mapping Different Approaches - 2009.pdf (Обект application/pdf). *Wiley Interscience*, *52*, 38–52. doi:10.1002/sd.244

Marchand, F., Debruyne, L., Triste, L., Gerrard, C., Padel, S., & Lauwers, L. (2014). Key characteristics for tool choice in indicator-based sustainability. *Ecology and Society*, *19*(3).

Ness, B., Urbel-Piirsalu, E., Anderberg, S., & Olsson, L. (2007). Categorising tools for sustainability assessment. *Ecological Economics*, *60*(3), 498–508. doi:10.1016/j.ecolecon.2006.07.023

Pope, J., Annandale, D., & Morrison-Saunders, A. (2004). Conceptualising sustainability assessment. *Environmental Impact Assessment Review*, *24*(6), 595–616. doi:10.1016/j.eiar.2004.03.001

Pope, J., & Morrison-Saunders, A. (2013). Pluralism in practice. In A. J. Bond, A. Morrison-Saunders, & R. Howitt (Eds.), *Sustainability assessment: pluralism, practice and progress* (pp. 100–114). Routledge.

Poppe, K., & van Asseldonk, M. (2015). *EIP-AGRI focus group benchmarking of farm productivity and sustainability performance: How can farmers and advisers use benchmarking data and process to improve productivity and sustainability performance?*

Riley, J. (2001). Indicator quality for assessment of impact of multidisciplinary systems. *Agriculture, Ecosystems and Environment*, *87*(2), 121–128. doi:10.1016/S0167-8809(01)00272-9

Schindler, J., Graef, F., & K??nig, H. J. (2015). Methods to assess farming sustainability in developing countries. A review. *Agronomy for Sustainable Development*, *35*(3), 1043–1057. doi:10.1007/s13593-015-0305-2

Steunpunt Duurzame Landbouw. (2006). *Erven van de toekomst*. Melle: Instituut voor Landbouw- en Visserijonderzoek (ILVO). Steunpunt Duurzame Landbouw (Stedula).

UNAIDS. (2010). *An introduction to indicators.* UNAIDS.

Van Passel, S., & Meul, M. (2010). Multilevel sustainability assessment of farming systems : a practical, (July), 791–800.

Van Passel, S., Nevens, F., Mathijs, E., & Van Huylenbroeck, G. (2007). Measuring farm sustainability and explaining differences in sustainable efficiency. *Ecological Economics*, *62*(1), 149–161. doi:10.1016/j.ecolecon.2006.06.008

Waas, T., Hugé, J., Verbruggen, A., & Wright, T. (2011). Sustainable Development: A Bird's Eye View. *Sustainability*, *3*(12), 1637–1661. doi:10.3390/su3101637

Weaver, P. M., & Rotmans, J. Working papers 1: Integrated sustainability assessment: What? Why? How? (2006).