



HAL
open science

Environmental sustainability assessment in agricultural systems: A conceptual and methodological review

Emma Soulé, Philippe Michonneau, Nadia Michel, Christian Bockstaller

► To cite this version:

Emma Soulé, Philippe Michonneau, Nadia Michel, Christian Bockstaller. Environmental sustainability assessment in agricultural systems: A conceptual and methodological review. *Journal of Cleaner Production*, 2021, 325, pp.129291. 10.1016/j.jclepro.2021.129291 . hal-03423461

HAL Id: hal-03423461

<https://hal.science/hal-03423461v1>

Submitted on 17 Oct 2023

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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Environmental sustainability assessment in agricultural systems: a conceptual and methodological review.

Emma Soulé^{a,b}, Philippe Michonneau^b, Nadia Michel^c, Christian Bockstaller^a

a Université de Lorraine, INRAE, LAE F68000 Colmar, France

b SCARA, F10700 Arcis sur Aube, France

c Université de Lorraine, INRAE, LAE F54500 Nancy, France

* Corresponding author: emma.soule@inrae.fr, INRAE LAE, 28 rue de Herrlisheim, 68000 Colmar, France; +33389224985

Acknowledgements

This study is part of a PhD research project supported by a grant (n°2019/0020) overseen by the French National Research Technology Agency (ANRT). The authors would like to thank Philippe Girardin for having initiated this research 15 years ago; Jérôme Rosnoblet and Chloé Schneller for their early contributions to the INDIC[®] database; Pauline Feschet and Emilie Thomas-Delille for the INDIC[®] database; Maria Lampridi (IBO, University of Thessaloniki) for the literature database; Allan Buckwell for the file of methods supporting his report; the members of the PhD steering committee, Frédérique Angevin, Philippe Jeanneret, Nicolas Munier-Jolain, Olivier Therond and Clotilde Toqué for their advice, and Jean Villerd for statistical support.

CRedit authorship contribution statement

Christian Bockstaller and Emma Soulé: Conceptualization, Investigation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Nadia Michel and Philippe Michonneau:** Conceptualization, Writing - review & editing, Visualization.

1 Wordcount: 10312

2 Abstract

3 As the sustainable development movement gains traction, many initiatives have aimed to provide
4 support with sustainability assessment methods using a loosely structured set of indicators, resulting in
5 an explosive development of such indicators. Several reviews have sought to gain a clearer
6 understanding of this “indicator zoo” by comparing assessment methods. However, most reviews
7 covered few methods and mainly focused on describing and evaluating them to help users select the
8 one that suited their needs. What is still lacking is an in-depth analysis of the conceptual and
9 methodological framework behind the sustainability assessment methods in order to provide
10 recommendations for assessment method developers and identify research gaps. To fill this gap, we
11 conducted a conceptual and methodological review on a sample of 262 studies covering the worldwide
12 agricultural sector between 1993 and 2019. Because the subject is so vast and due to disciplinary
13 barriers, we restricted this review to studies with an environmental dimension but did not set any
14 geographic limitations (the studies cover both temperate and tropical zones). The initial results show a
15 need for clarity on terminology. Methods should also explicitly specify the purpose, target users and
16 temporal scale. Additional efforts must be made in selecting indicators on the causal chain at emission
17 level or those that pertain to impacts, such as in life cycle assessment. There are additional research
18 gaps when more than a simple sum of scores is required to aggregate indicators, or when dealing with
19 upscaling or spatialization issues. Dynamic assessment of resilience or robustness with multiple
20 criteria also remains a challenge. Environmental sustainability assessment methods will also need to
21 integrate ecosystem services and emerging environmental issues such as pollution from antibiotics or
22 microplastics. Finally, a similar study on assessment methods should be conducted with a focus on
23 social and economic dimensions.

24

25 Keywords: indicator, multi-criteria assessment, aggregation, reference value, life cycle assessment

26

27

28 **1. Introduction**

29 Agriculture remains the backbone of world development, providing humans with food and other goods
30 such as bioenergy, fibres and more (Godfray et al., 2010). However, it is also a major cause of the
31 Earth’s planetary boundaries being pushed beyond their limits, which confirms the tremendous
32 pressure that modern agriculture exerts on sustainability (Campbell et al., 2017). Thus, any credible
33 roadmap to improve the situation on a global level must drastically reduce agriculture’s environmental
34 footprint. While stakeholders around the planet – from farmers to international decision-makers –
35 unanimously agree about the general goals of sustainability, ideas on how these goals should be
36 reached are fiercely debated (Binder et al., 2010). Organic farming is a good illustration of this lack of
37 consensus: it is promoted by environmentalists but is challenged as a solution for sustainable
38 agriculture because of lower yields (Ponisio et al., 2015). Despite disagreement when it comes to
39 achieving sustainability, most stakeholders recognize the need for assessment methods to support
40 attempts to identify and implement new measures (Pope et al., 2004), to characterize the current
41 situation (Wang et al., 2009), and to gauge progress that must be made to reach the end goal (Bell and
42 Morse, 2001). The challenging task of assessing a concept as complex as sustainability using metrics
43 in an operational way, i.e. with data available to stakeholders, has quickly led to the use of indicators
44 instead of directly measuring sustainability itself (Mitchell et al., 1995). As early as 2001, Riley
45 (2001) spoke of an “indicator explosion” that has not stopped since, as pointed out by Pintér et al.
46 (2012): “we continue moving toward an ‘indicator zoo’, characterized by a multitude of approaches”.

47 It is no surprise that various reviews have been published in response to this explosive development to
48 characterize and compare both indicators and sustainability assessment methods (SAMs) and help
49 potential users pick their way through this “myriad of methods” (Coteur et al., 2019). A highly cited
50 review by Singh et al. (2012, 2009) described 41 different methodologies and indices implemented

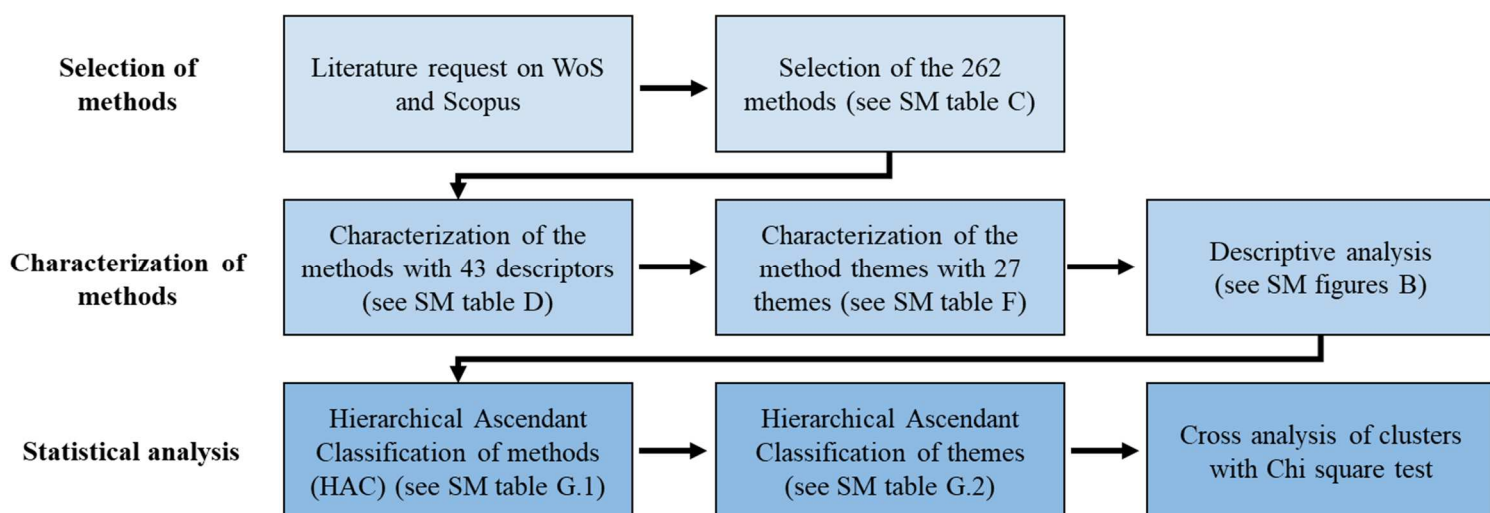
51 mainly at the national and international levels, such as the “Human Development Index” or the
52 “Compass Index of Sustainability”, and categorized them according an updated version of a
53 framework proposed by Ness et al. (2007). Several reviews and comparison studies that were more
54 relevant to the agriculture sector have been published after a pioneer paper by van Der Werf and Petit
55 (2002), which mainly dealt with a limited number of methods. Recently, Lampridi et al. (2019)
56 performed a small meta-review of 16 review articles on SAMs and indicators published since 2005,
57 and then reviewed 38 methods. However, they overlooked several review articles such as those by
58 Schader et al. (2014), de Olde et al. (2016) and Eichler Inwood et al. (2018). These authors began with
59 large samples (35, 48 and 48 methods, respectively), but then conducted more in-depth analyses on
60 restricted samples (6, 4 and 9 methods, respectively). Although not published in a peer-reviewed
61 article, the study by Wustenberghs et al. (2015) also deserves attention: it analysed 53 methods
62 selected from a list of 170. Lampridi et al. (2019) analysed descriptors and evaluation criteria which
63 were used in reviews of SAMs. Most provided general information on the methods (e.g. publication
64 year, method purpose, target users, geographic scope, level of assessment etc.) while some authors
65 focused on specific points. Schader et al. (2014) looked at assessment precision; de Olde et al. (2016)
66 covered operational aspects such as time requirement data availability; Eichler Inwood et al. (2018)
67 examined spatial and temporal boundaries, dimensions, themes, indicators and methodological
68 aspects; and Wustenberghs et al. (2015) focused on stakeholder participation and indicator
69 characteristics. Many reviews addressed a specific indicator-focused theme such as pesticides
70 (Bockstaller et al., 2009; Feola et al., 2011). Rasmussen et al. (2017) reviewed 84 studies and
71 identified the “top three” themes covered: soil quality, greenhouse gases (GHG) and soil erosion.

72 This literature survey showed that most reviews dealt with a limited number of methods or case
73 studies and focused on the most useful considerations for potential users: general descriptive aspects,
74 operational issues or evaluation criteria of advantages and drawbacks (e.g. transparency, sensitivity,
75 data reliability). However, method design with concepts, assumptions and underlying methodologies
76 could have been more deeply analysed to gain generic knowledge for a state of art and identify
77 research gaps. For example, at the conceptual level, most studies listed only those aspects
78 corresponding to the three basic sustainability dimensions (economic, environmental and social)
79 covered by a method. Little was mentioned regarding their positioning in terms of three possible
80 sustainability orientations: management-oriented, goal-oriented or property-oriented (Alkan Olsson et
81 al., 2009; Hansen, 1996). Indicators are often described only in terms of a quantitative versus
82 qualitative typology. They could have been analysed in much greater detail, such as by considering
83 their positioning on the causal chain or their typology according to Bockstaller et al. (2015), who
84 distinguished causal indicators from effect indicators. Causal indicators are based on single or simple
85 combinations of variables, while effect indicators are derived either from model outputs or
86 measurements. Such studies could be used to draw up methodological recommendations for
87 assessment method developers and to identify research gaps.

88 To fill this gap, we conducted a conceptual and methodological review on a huge sample of 262
89 studies. This sample encompassed i) **methodological frameworks** with general characteristics, such
90 as guidelines for indicator selection used to develop SAMs (a typical example of such a framework is
91 the life cycle assessment – LCA (Haas et al., 2000; Repar et al., 2017)); ii) **dedicated methods**, which
92 refer to generic SAMs based on a list of sustainability indicators that are more or less structured within
93 a conceptual framework; these methods may or may not be aggregated at different levels and are
94 meant to be implemented by end users; and iii) **case studies** where authors applied a list of indicators
95 selected specifically for their study. Due to disciplinary barriers and because we considered the subject
96 to be sufficiently vast, we restricted this review to only studies with an environmental dimension
97 (hereafter ESAM) but did not set geographic limitations, with studies covering both temperate and
98 tropical zones. This review aims to produce a state of art of the various published methods to identify
99 research gaps in the field of sustainability assessment and provide ESAM developers with avenues for
100 improvement. Key terms are defined in a glossary in Table A in the supplementary materials.

101 **2. Materials and methods**

102 Figure 1 shows the different steps of the methodological path from the selection of the methods to the
103 statistical analysis.



104 Figure 1: Flowchart of the methodological path followed, including references to supplementary
 105 materials.

106

107

2.1. Selection of methods

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

The review process followed the main steps of the framework for systematic reviews in agricultural science (Koutsos et al., 2019) derived from the most widely used and more formal PRISMA framework developed by Moher et al. (2010). Our inventory covered the period between 1993 and 2019 and included scientific literature and grey literature. We ran the equation of Rasmussen et al. (2017) on Web of Science and the equations of Lampridi et al. (2019) on Scopus, applied to the study period (see equations in Figure B.1 in the supplementary materials). This database query was completed with ESAM from the inventories of Buckwell et al. (2014), de Olde et al. (2016), and Wustenberghs et al. (2015b) and our research group, who built the INDIC[®] database (GIS GC HP2E, 2018). These four reviews were carried out using a snowballing approach, where the reference list and citations of relevant papers were reviewed to identify new papers. Two authors reviewed the papers; one examined the entire set of papers and the second author reviewed at least half of them to reach a consensus. As noted by Levitan (2000) with regard to the pesticide risk indicator, several ESAM were published only in grey literature in reports, or more recently on websites. The whole inventory resulted in a dataset of 4,523 dedicated methods and case studies. We applied a set of elimination criteria (see Table 1) to eliminate any methods which did not have an environmental dimension or pertain to the agricultural sector, such as those focusing only on landscape or forestry management. The last criterion for articles published about a simple application of a dedicated method or methodological framework such as LCA helped eliminate the voluminous literature on LCA implementation. Only papers with methodological innovations were retained (e.g. a pioneering paper on adapting LCA to agriculture or a specific agricultural sector). After applying the elimination criteria, we obtained an inventory of 262 ESAM and case studies (see Table C in the supplementary materials for the full list of references). As recommended by the PRISMA framework (Moher et al., 2010), the different steps are summed up in Figure B.2.

Table 1. List of elimination criteria.

Elimination criteria	Explanation
Duplicate	Duplicates appearing due to cross-referencing of bibliographical sources.
Beyond the scope	Method assessing an area other than environmental sustainability and agricultural sector.
Monothematic	Method evaluating only one (or sometimes two) themes such as water quality.

Lack of detailed information	Indicators structuring the assessment method were not described.
Inaccessible information	Only the abstract, not the full paper, was accessible.
Old version of a selected method	Only the latest version of a method was inventoried.
Discontinued method	The method is no longer used.
Application of a dedicated method or method derived from a framework	The article described an application of a dedicated method or a methodological framework already described in another publication.

133

134

2.2. Selection of descriptors to characterize methods

135

When labelling methods, we decided to use the terminology from de Olde et al. (2016) – “dimension/theme/subtheme/indicator” – which is widely used and easy to understand. These authors noted the wide range of terms in use; for example, “indicator” is interchangeable with “metric” (Sikdar, 2003). Likewise, we use the term “method”, which some authors refer to as “tool”, “methodology”, “approach” etc. (Sala et al., 2015).

136

137

138

139

140

141

We defined a set of 43 descriptors to characterize the ESAM (Table D in the supplementary materials), broken down into i) general descriptors, ii) descriptors covering the conceptual sustainability framework based on various themes, iii) descriptors for the indicator type, structure and design, and iv) descriptors for the system definition including scale or system boundary issues. We selected general descriptors such as method origin, geographic scope, purpose, target users, number of dimensions, themes and indicators, etc. based on the aforementioned review of Lampridi et al. (2019) and Wustenberghs et al. (2015). The various terminology used to describe the different method components (see previous paragraph) was inventoried to characterize the conceptual frameworks. We also added descriptors such as the sustainability definition orientation chosen by the method designer based on different main principles (Hansen, 1996). This orientation depends on the formulation of the item (themes, criterion etc.). The framework may be i) management oriented when items are based on a management requirement (e.g. low pesticide use); ii) goal oriented when there is a direct link to an environmental issue (e.g. water quality or “pesticide in water”; Alkan Olsson et al., 2009); or iii) property oriented (e.g. productivity, López-Ridaura et al., 2005). It is important to consider the position of an indicator on the causal chain during analysis, as this determines the indicator’s ability to predict an impact (Bockstaller et al., 2015; Repar et al., 2017). The typology in Bockstaller et al. (2015), which combines the indicator’s position on the causal chain (causal versus effect), and the way indicator value is determined (predicted vs. measured) was also used to distinguish:

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

- causal indicators based on a single variable or simple combination of variables (mainly management), called “means-based indicators” by van der Werf and Petit (2001);
- predictive effect indicators derived from an operational and easy-to-implement model output or from a complex research-based model;
- measured effect indicators obtained from measuring an effect.

159

160

161

162

163

164

In addition to these conceptual descriptors, we used several others to describe methodological aspects such as the determination of indicators and reference values as well as aggregation. It is important to link an indicator to a reference value because the difference between the indicator raw value and the reference value is more informative than the raw value alone (Acosta-Alba and van der Werf, 2011). Aggregation can be problematic when it comes to trying to strike a balance between simplifying data and ensuring sufficient details and transparency (Tzilivakis et al., 2016). Opponents of aggregation often cite loss of information, subjectivity in the procedure and weighting issues as key problems. However, Bockstaller et al. (2008) argued that such issues can be overcome because non-aggregated indicators are often needed to analyse systems while aggregated indicators are used to compare systems and provide conclusions. Transparency can be determined by the degree of explicitness in the weighting procedure (Greco et al., 2019). Interaction between indicators is addressed qualitatively to know whether dependency exists between indicators in the calculation or interpretation phase (de Olde et al., 2016).

165

166

167

168

169

170

171

172

177 Lastly, we selected descriptors related to the system description. Temporal and spatial scale issues are
 178 important and were described more precisely than with a simple descriptor as in Wustenberghs et al.
 179 (2015b): details included their different components, resolution (the smallest grain), extent (the whole
 180 area or period considered) (Purtauf et al., 2005) and upscaling issues (Dalgaard et al., 2003). For
 181 temporal issues, descriptors were also derived from the typology in Ness et al. (2007): either
 182 retrospective, based on a real past situation (*ex post*) or prospective, based on a possible situation (*ex*
 183 *ante*). System boundaries, and more precisely the degree of coverage of a product’s life cycle, was
 184 also addressed. LCA not only covers direct effects linked to production systems (as in “conventional”
 185 assessments), but also deals with indirect effects that occur downstream due to inputs as well as
 186 upstream due to product use and waste (Repar et al., 2017). Nevertheless, partial LCA on some themes
 187 such as energy consumption or GHG emissions must also be considered (Bockstaller et al., 2015).
 188

189 2.3. Characterization of themes addressed by the methods

190 To take the analysis one step further than most reviews, which only deal with the dimension level, we
 191 also characterized the major themes. An indicator-level analysis would have been beyond the scope of
 192 such article due to the much higher diversity than at theme level (see, for example, reviews on
 193 nitrogen loss by Bockstaller et al. (2015); Buczko and Kuchenbuch (2010)). To cover the broad range
 194 of environmental dimensions, we selected 67 themes, among which are classic themes in the literature
 195 such as water and soil quality (Wustenberghs et al., 2015) and others based on our expertise. We
 196 divided them into several categories based on the causal chain:

- 197 - causal: landscape elements, management practices (e.g. tillage, crop inputs, livestock);
- 198 - system functioning: having to do with efficiency or resource/nutrient balances (e.g. water,
 199 nitrogen);
- 200 - agroecosystem services such as carbon storage and pest regulation that can reduce impacts or
 201 input use (Therond et al., 2017);
- 202 - impacts: derived from LCA or other assessments (e.g. soil quality, erosion; see Repar et al.,
 203 2017; Sadok et al. 2009);
- 204 - systemic properties such as autonomy (López-Ridaura et al., 2005; Zahm et al., 2015).

205 Our analysis focused mainly on the 17 most represented themes among the reviewed ESAM, which
 206 occurred in at least 20% of studies.
 207

208 2.4. Statistical analysis

209 Following a descriptive analysis, we performed a multivariate analysis to look for trends and clusters.
 210 We performed a first multiple correspondence analysis (MCA) for the 262 ESAM on 12 descriptors
 211 that we felt were most relevant and a second MCA on 17 themes (see Table 2) using RStudio
 212 version 1.3.959 and the FactoShiny package (Husson et al., 2010). Next, we performed a hierarchical
 213 ascendant classification (HAC) to obtain several clusters using the FactoMineR package. Finally, we
 214 ran a Chi-Square test to cross-check the HACs to determine whether the ESAM and theme clusters
 215 were independent or dependent.
 216

217 Table 2: List of the selected descriptors and themes for the hierarchical ascendant classification.
 218

Method descriptors	Themes	
Age	Biodiversity	Crop management
Literature	Fertilizer management	Species diversity
Purpose	Pesticide management	Water management
Target user	GHG emissions	Crop diversity
Production	Soil erosion	NO ₃ leaching
Orientation	N balance	
Position	Soil organic matter	
Indicator type	Energy consumption	
Aggregation method	Water quantity	

Assessment type	Land use
LCA methodology	Water quality
Spatial resolution	Ecosystem diversity

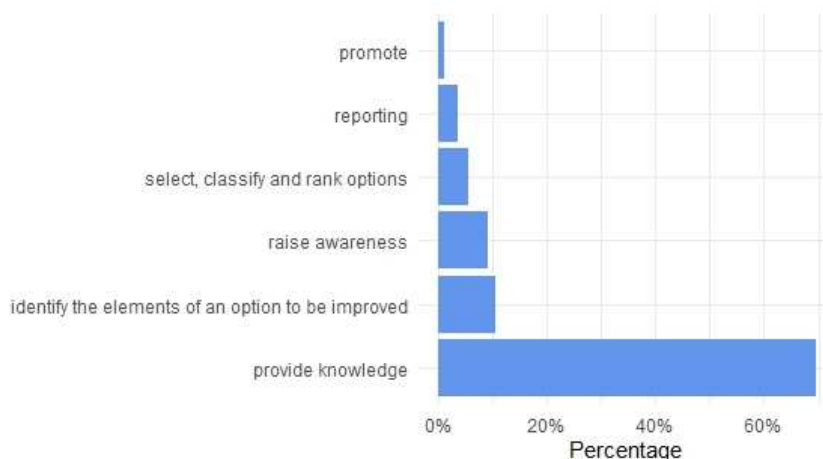
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249

3. Results

3.1. Characterization of methods

3.1.1. General information on methods

The research equations described in Lampridi and Rasmussen logically resulted in the most ESAM, followed by the INDIC® database (Figure B.3a). With regard to the chronological distribution between ESAM sources (Figure B.3b), we observed that the INDIC® database provided the most methods for the first half of the review period; the equation research resulted in more methods for the second half. Most of the methods in our inventory came from scientific literature (81%). A total of 60% of studies dealt with a “dedicated method” while “case studies” represented only 37%; the remainder (3%) were “methodological frameworks”. Our ESAM sample covered 60 different countries in all. The major contributing countries were France (12%), Italy (8%), Spain (5%), China (5%) and Brazil (4%) (Figures B.4a and B.4b). Several ESAM were also developed in countries in the Global South, including India, Pakistan and Iran. For some countries like France, the United States, Australia and Canada, between 55% and 60% of the methods published in each country came from the grey literature (Figure B.4b). One explanation could be that a mastery of French and English languages by authors favoured the selection of such literature, although some authors did read in other languages (German, Spanish, Italian). The yearly number of publications since the 1990s steadily increased until 2014 before dropping off; a similar trend was also found by Rasmussen et al. (2017) (Figure B.5). The proportion of methods published in the grey literature did not show any clear temporal trend. The drop in publications after 2014 occurred in both types of literature. One possible explanation could be the emergence of the ecosystem services concept and a shift away from developing SAMs to those dealing with ecosystem services. In most of the ESAM, a specific production sector was not specified, so we grouped them together under “various sectors”. This category is followed by “field crops” and “crop productions”, “mixed farming” and “dairy farming” (Figure B.6). As noted by Bockstaller et al. (2015), the purpose of the assessment method should be specified because it can guide the selection of an ESAM. Nevertheless, we found that nearly three-quarters of methods sought to “provide knowledge”, and so we attributed this category to the method when no purpose was specified (Figure 2). This purpose far outranked the next categories and was followed by “identify the elements of an option to be improved” (11%) and “raise awareness” (10%).



250
251
252
253
254
255

Figure 2: Percentage of articles (n=262) according to the main purpose of the methods.

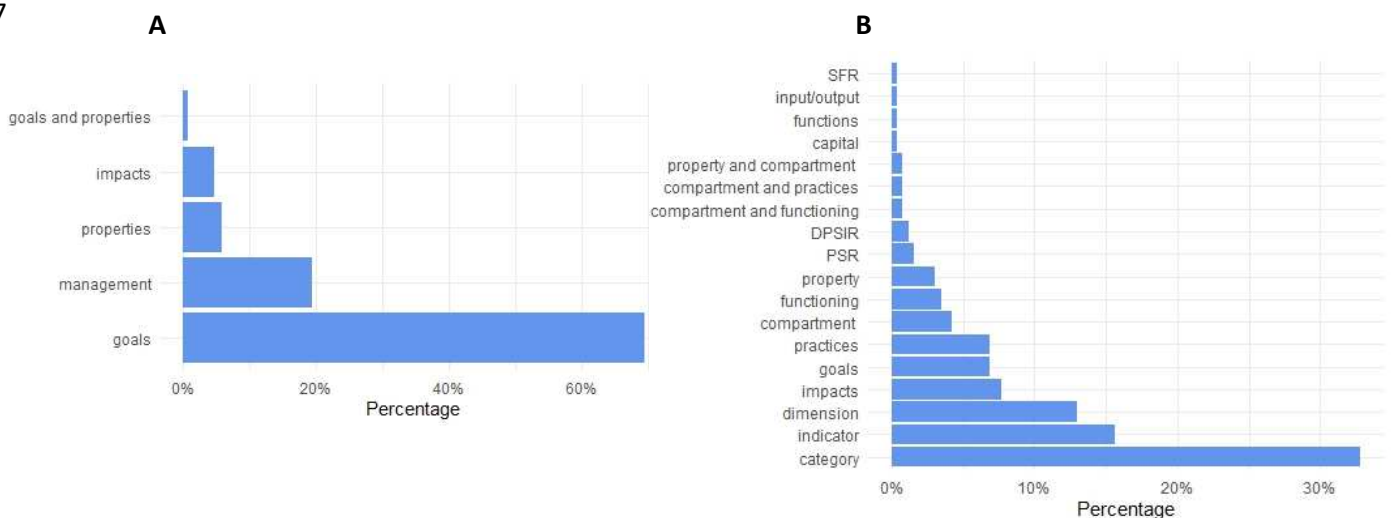
3.1.2. Conceptual framework

Figure 3A shows that most of the methods are goal-oriented (67%), followed by management-oriented methods. Although proponents of a property-oriented approach claim that a systemic approach is better addressed by a property orientation than by management or goal orientations (e.g. Bossel et al.,

256 2000), few methods (6%) adopted such an approach. A goal-oriented approach is both more pragmatic
 257 and much easier to understand than a property-oriented approach. However, goal-oriented approaches
 258 were explicitly defined only for 6% of the ESAM while 84% of the ESAM were poorly defined and
 259 10% were defined in a limited way.

260 Themes are mostly organized into “categories” that encompassed a combination of goals or
 261 environmental compartments, management practices or even properties (e.g. autonomy) (Figure 3B).
 262 The next items are “indicators” (15%) and “dimensions” (13%), which indicate whether a method is
 263 based on a unique indicator list without themes, or if indicators are only grouped within the three or
 264 more sustainability dimensions. Many authors did not use well-known frameworks such as DPSIR or
 265 PSR (Niemeijer and de Groot, 2008). One explanation is that there are few methods dealing with the
 266 country spatial level (see Figure 5B).

267



268 Figure 3: Percentage of articles (n=262) according to A) method orientation and B) criteria
 269 organization. SFR: structure function response framework, see Wang et al. (2009); DPSIR: driving-
 270 force pressure state impact response; PSR: pressure state response (Niemeijer and de Groot, 2008).

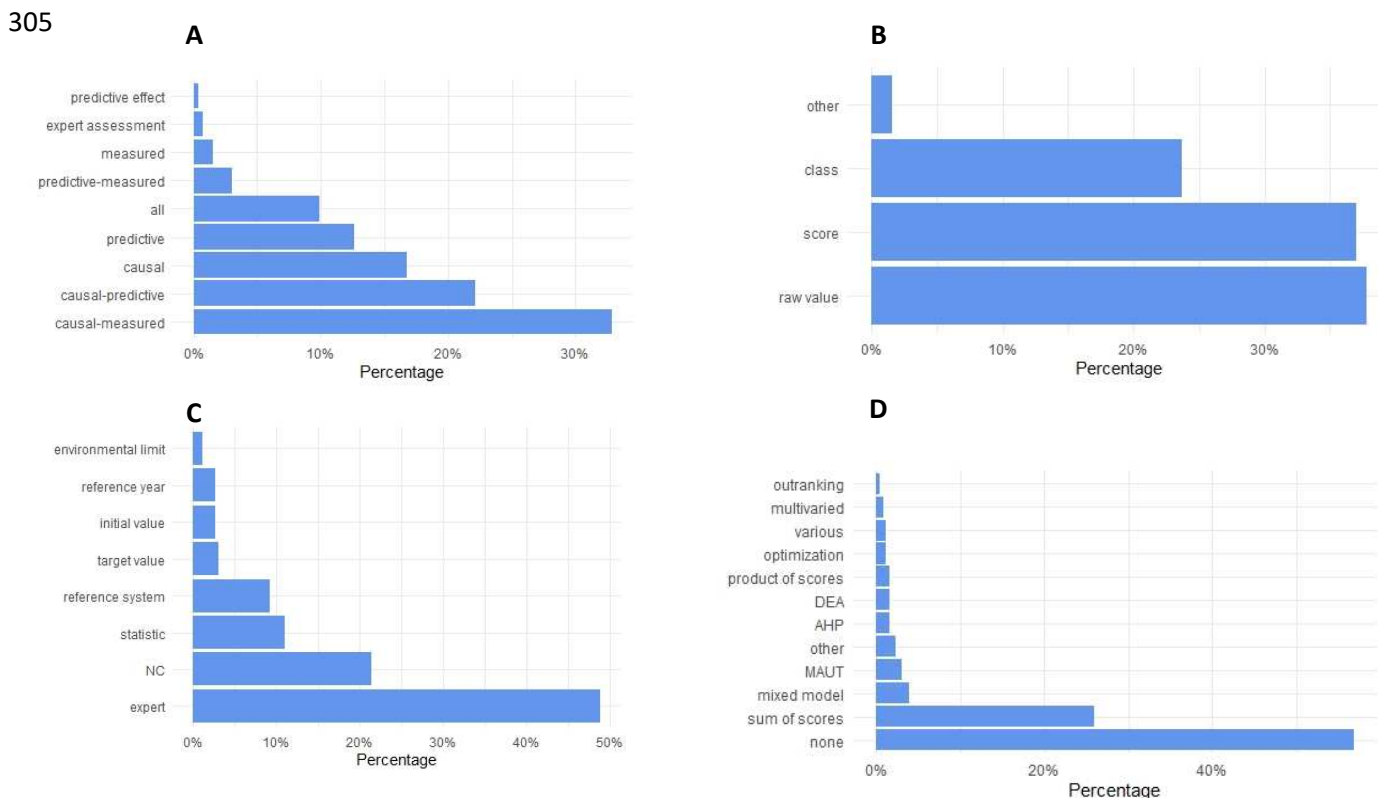
271 The terminology used by authors to present their frameworks provided additional information (Figure
 272 B.7). De Olde et al. (2016) noted the variety in the terminology, and our study more than confirmed
 273 this observation: we inventoried more than 100 different word combinations. The main term
 274 “indicator” was used alone in 30% of articles, reflecting methods based on a list of indicators, and was
 275 followed by the terms “dimension – indicator” (9%). This is where the consistency ended and the
 276 confusion began. The terms “criterion – indicator” or “theme – indicator” account for only 3% of
 277 articles, while the terminology proposed by de Olde et al. (2016) was only used by Lebacqz et al.
 278 (2013). We also noticed that some ESAM used the same terms but transposed them, such as “attribute
 279 – indicator” and “indicator – attribute” (e.g. Duncombe-Wall, 1999; Walker, 2002). The term
 280 “indicator” in the latter pair of terms is not employed as a metric to assess a farming system
 281 characteristic but rather refers to the conceptual item guiding the selection of a metric.

282 3.1.3. Indicators

283 Most methods were based on indicators calculated only with management data (Figure B.8). More
 284 specifically, 15% ESAM were designed with an indicator to assess the management strategy and state
 285 of the environment; in most cases, the environmental state indicator was based on an easily determined
 286 measurement such as soil observation (as in Candelaria-Martínez et al., 2014; Gomez-Limon et al.,
 287 2010; Yegbemey et al., 2014). A total of 13% ESAM were based on management strategy alone.
 288 Around 10% of the ESAM belonged to the category “management-emission-state”. Some were close
 289 to the first category when they included only one additional indicator assessing GHG emissions with a
 290 calculation method derived from operational IPCC Tier 1 or 2 calculations (IPCC, 2003), such as in El
 291 Chami and Daccache (2015); Kuneman and Fellus (2014); and Uhlman et al. (2016). The next
 292 category was “management-functioning” (7%). These were methods that used indicators based on
 293 management strategy and included one or two indicators calculating a nutrient or water balance or

294 efficiency, such as in Laurent et al., 2017 and Moraine et al., 2017. The likely explanation for why
 295 most methods used indicators assessing management strategies is that they are easier to implement
 296 when data are available compared to calculating or measuring emissions or the environmental state
 297 (Niemeijer and de Groot, 2008; Payraudeau and van Der Werf, 2005; Repar et al., 2017).

298 The predominance of management indicators is also reflected by a majority of the ESAM (70%) based
 299 on “causal indicators”, which may or may not be associated with measured effect or predictive effect
 300 indicators (Figure 4A). Only 12% of the ESAM used only predictive indicators derived from model
 301 simulation outputs or operational models, which in their simplest form can indicate an emission factor.
 302 Methods derived from the LCA framework belong to this group. Less than 5% included only
 303 measured indicators that can be easily understood. Generally speaking, measurements require more
 304 time and money (Bockstaller et al., 2015).



306 Figure 4: Percentage of articles (n=262) according to A) indicator typology (see Bockstaller et al.,
 307 2015), B) indicator setting, C) nature of the reference for indicators, and D) type of aggregation
 308 methods identified. NC: Not concerned; AHP: analytic hierarchy process, see Egea and Pérez y Pérez
 309 (2016); DEA: data envelopment analysis, see Ullah and Perret (2014); MAUT: multi-attribute utility
 310 theory, see Van Cauwenbergh et al. (2007).

311 With regard to the general type of indicators, 78% of the ESAM were based on quantitative indicators
 312 while only 6% were strictly qualitative and 16% included both types. Figures 4B, 4C and 4D provide
 313 more information on indicator formatting. Most ESAM were designed with indicators expressed by a
 314 raw value of the measurement of a calculation or by a score, i.e. converting a raw value or category
 315 into a numerical value on different scales (Bockstaller et al., 2008). About 22% of the ESAM used
 316 qualitative indicators expressed by class. A large proportion of the ESAM (83%) linked a reference
 317 value to the indicators, most determined by experts (i.e. by the authors themselves, or in 128 of 218
 318 the ESAM, via a participatory approach). Among methods with a reference value, 67% used an
 319 absolute value, but some were based on scientific data (environmental limit or target value, Figure 4C)
 320 while 33% of the ESAM used relative reference values. In this case, indicator values were expressed
 321 in relation to the value of a given year or the initial year of study (5%), to the value calculated for a
 322 reference system (e.g. conventional system) or to the calculation of a statistic (average, maximum or
 323 minimum value; 8.4%). This last class of reference value was often related to a normalization step

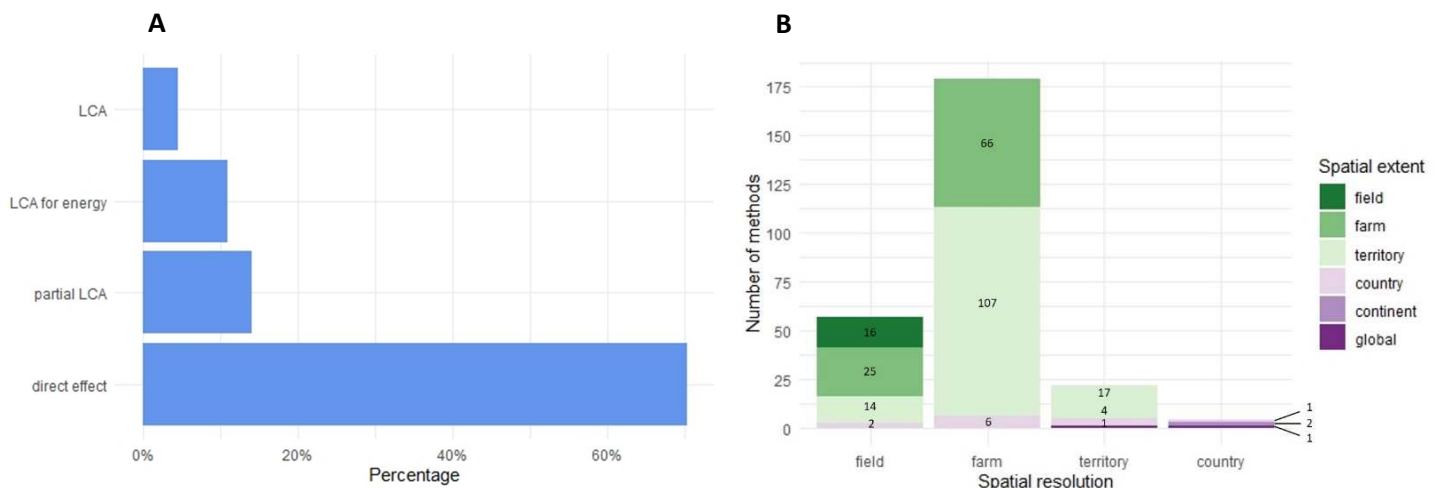
324 before any aggregation procedure. A total of 16% of the ESAM involved aggregation by dimension or
 325 global sustainability, while aggregation by criterion was found in less than 10%. As Figure 4D shows,
 326 almost half the ESAM did not include an aggregation procedure and about 26% consisted of a sum of
 327 scores that require a normalization procedure, raising methodological discussion on subjectivity
 328 (Pollesch and Dale, 2015). In contrast, several well-known aggregation methods with a solid
 329 methodological foundation (Greco et al., 2019), such as outranking methods or mixed models, were
 330 used by few authors. Among the ESAM using aggregated indicators, 60% were developed with an
 331 explicit weighting procedure, whereas 9% were not totally transparent and 31% avoided the problem
 332 by not using weighting.

333 3.1.4. Systems

334 About 70% of the ESAM dealt only with direct effects (Figure 5A), which is unsurprising since Repar
 335 et al. (2017) called them “conventional assessment methods”. Only 5% of methods included LCA
 336 methods, but this is because our review did not include the enormous amount of literature on LCA
 337 implementation. As pointed out by Bockstaller et al. (2015), “conventional assessment methods” may
 338 include a partial LCA approach: in our review, this applied to 14% of the ESAM for energy
 339 consumption when they included indirect consumption (Castoldi and Bechini, 2010; Fernandes and
 340 Woodhouse, 2008) and 11% for several themes such as GHG (Hanegraaf et al., 1998; Nasca et al.,
 341 2006).

342 Figure 5B shows that the main spatial scale combination covered by the ESAM was calculation at
 343 farm level, either for a group of farms in a given territory (41%), or for a single farm (25%). This is
 344 not surprising as the farm is the central component decision unit in an agricultural system. Higher
 345 resolution at field and lower resolution at territory or country were less covered. The temporal
 346 resolution of assessment was not explicit in many methods (24%), but was implicit in the calendar
 347 year or the cropping period (from harvest to harvest). In 75% of the ESAM, the assessment was
 348 carried out for one year; in the modelling study by Sands and Podmore (2000), calculations were
 349 performed at day resolution. Furthermore, 83% were *ex post* methods, 13% were *ex ante* and 4% were
 350 both.

351



352
 353 Figure 5: Percentage of articles (n=262) according to A) the degree of use of the LCA methodology,
 354 and B) spatial resolution and extent chosen by the method.

355 3.1.5. Cluster analysis of methods

356 The 262 methods were clustered hierarchically into four clusters. We analysed the way in which
 357 characteristics were distributed within each cluster in order to label them: M1, M2, M3 and M4 with
 358 62, 124, 59 and 17 methods, respectively (Table G.1 in the supplementary materials). We used cluster
 359 M2, which had the most methods, to identify dominant characteristics and then noted specific

360 characteristics for the other clusters (Figure 6). The dominant characteristics are: i) a published
 361 method assessing sustainability at farm level, ii) providing knowledge to farmers, iii) aimed at various
 362 sectors with a goal-oriented framework, iv) delivering non-aggregated causal and measured
 363 management indicators (see Section 2.2.), and v) assessing direct effects. M1 is relatively similar to
 364 M2, although most of the ESAM were only published in the grey literature for farmers and had only
 365 causal indicators. Some of these methods aimed to increase farmers’ awareness of sustainability
 366 themes, i.e. to provide them with general information to make them aware of a problem and encourage
 367 them to rethink their management practices. M3 and M4 are considerably different. Methods in M3
 368 cover field crops and are aimed at *ex ante* assessment at field level. When specified, the purpose is to
 369 “select, classify and rank options”, typically multi-criteria decision aid (MCA) problem. Indicators
 370 assess system functioning (balance, efficiency) and even emissions (see Table G.1 in the
 371 supplementary material), and some methods include predictive indicators. M4 covers LCA-based
 372 methods (Moreau, 2016; Wang et al., 2019), which are impact-oriented methods with predictive
 373 indicators, including an impact assessment model as specified in LCA. Although M4 covers mainly
 374 field crops, this does not reflect the reality of LCA that was applied in many agricultural sectors but
 375 rather is due to the restricted selection of papers with a methodological innovation.

Method clusters	Literature type	Purposes	Target users	Production sector	Orientation	Position on causal chain	Indicator type	Aggregation method	Assessment type	Level of LCA methodology	Spatial resolution
M2 (n=124)	Scientific	Provide	Researchers	Various sectors	Goal	Management	Causal measured	No aggregation	Ex post	Direct effect	Farm
M1 (n=62)	Grey	Raise	Farmers	Various sectors	Goal	Management	Causal	Sum of scores	Ex post	Direct effect	Farm
M3 (n=59)	Scientific	Select	Researchers	Field crops	Goal	Functioning	Causal predictive	No aggregation	Ex ante	Partial LCA	Field
M4 (n=17)	Scientific	Provide	Researchers	Field crops	Impact	Impacts	Predictive	Sum of scores	Ex post	LCA	Field

376 Figure 6: Illustration of hierarchical ascendant classification analysis characterizing the methods
 377 according to main (in grey) and specific (in colour) characteristics. Pale-coloured specific
 378 characteristics are not specific predominant characteristics but are important enough to be represented.
 379 For the purposes, “Provide” refers to provide knowledge, “Raise” to raise awareness and “Select” to
 380 select, classify and rank options.

381

382 3.2. Themes

383 3.2.1. Information on themes

384 Among the 262 selected ESAM, 10 themes stood out as they were found in more than 27% of the
 385 articles (Table 3). Biodiversity seems to be a key theme since it appears in almost half the articles,
 386 unlike themes related to ecosystem services. Indeed, only 5% of the methods explicitly mentioned at
 387 least one ecosystem service such as C storage (5%), food production (3%) or soil fertility services
 388 (2%). In Rasmussen et al. (2017), the five main indicators were: soil fertility, GHG emissions, soil
 389 erosion, land cover composition and nitrogen balance. In fact, five themes from Table 3 can be linked
 390 to indicators in Rasmussen et al. (2017): GHG (ranked second in both our study and in Rasmussen),
 391 fertilizer management (four and tenth, respectively), soil erosion (fifth and third, respectively), nutrient
 392 efficiency/balance (sixth in both studies) and soil organic matter (ninth and seventh, respectively).
 393 Despite these similarities, the ranking in our study was more diverse than in Rasmussen’s study, who
 394 included indicators covering the same themes. Thus, two indicators of physical soil conditions and one
 395 on soil fertility, as well as two indicators on nutrient balances, ranked in the top ten in Rasmussen et
 396 al. (2017).

397 Table 3: Frequency in percentage of the ten main themes found in the reviewed articles.

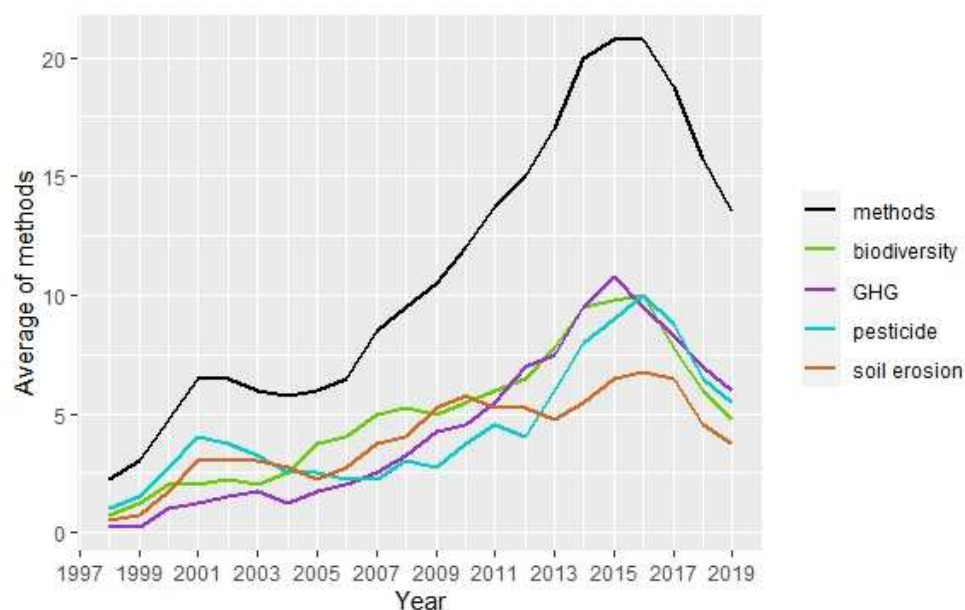
Themes	Reviewed articles
Biodiversity	44%

GHG emissions	41%
Pesticide management	40%
Fertilizer management	36%
Soil erosion	35%
Nutrient efficiency/balance	31%
Water quantity	29%
Energy consumption	29%
Soil organic matter	28%
Water quality	27%

398
399
400
401
402
403
404
405
406
407
408

3.2.2. Evolution of the themes covered since 1998

We calculated sliding averages over four years to smooth the evolution. Due to lack of data for the years 1994 and 1996, the sliding averages started in 1998. Figure 7 shows the evolution of the inclusion in ESAM of the four main themes from Table 3 between 1998 and 2019. There was a peak between 2015 and 2016, immediately followed by a significant drop. The themes of soil erosion and GHG emissions showed the greatest contrast in their evolution. The GHG emissions theme increased sharply in the 2010s while soil erosion was covered more regularly across the period with a smaller rise before stagnating from 2010 onwards. Biodiversity increased steadily since its emergence. The pesticide theme declined slightly in the 2000s before rising sharply, especially in the 2010s.



409 Figure 7: Evolution of the inclusion of four main themes in the methods since 1998 (sliding average
410 over four years).

3.2.3. Cluster analysis of themes

411 The main themes addressed by the 262 ESAM were clustered hierarchically into four clusters: T1, T2,
412 T3 and T4 with 65, 101, 31 and 65 ESAM, respectively. The distribution was slightly more balanced
413 than for the methods. After analysing the distribution of each theme in the cluster (see Table G.2 in the
414 supplementary materials), we then characterized each cluster (Table 4). While T1, T3 and T4 could be
415 clearly labelled, T2 showed no clear trend. This could be because this cluster contains ESAM with
416 very different associated themes that could not be put in a differentiated cluster. T3 shows the broadest
417 range of themes, including both environmental and management themes, while T1 and T4 each had a
418 narrow focus on either management or environmental themes, respectively. Surprisingly, cluster T1,
419

420 which had two main management themes (pesticide and water management) did not include the
 421 absence of fertilizer management, which is more closely associated with nutrient emissions. Themes
 422 on ecosystem services or systemic properties did not influence the results because of their low
 423 representation.

424 Table 4. Hierarchical ascendant classification analysis of the themes based on four clusters with a
 425 description and the number of methods for each cluster.

Cluster	Description
T1 (n=65)	Methods with management themes such as pesticides, water management and crop diversity.
T2 (n=101)	No clear trend. Two themes – nitrogen balance and greenhouse gas emissions – were most frequent.
T3 (n=31)	Methods addressing a broad array of management and environmental themes.
T4 (n=65)	Methods based on limited environmental themes such as GHG emissions, biodiversity, soil erosion, energy consumption and water quality.

426

427 3.3. Cross-checking method and theme clusters

428 The Chi-Square test shows a very high significant dependency between method clusters and theme
 429 clusters ($p\text{-value} = 1.59 \cdot 10^{-9}$). While the M1 method cluster has an equal distribution among the theme
 430 clusters (Table 5), this is not the case for M4, which has more methods in T4. For M2 and M3, the
 431 main theme cluster is T2, which seems to present a high variety of theme combinations. However, the
 432 second most represented theme cluster is T1, characterized by management themes for M2 whereas
 433 M3 and M4 are both characterized by T4.

434

435 Table 5. Contingency table of the Chi-Square test on method and theme clusters.

436

	T1	T2	T3	T4
M1	14	16	18	14
M2	47	50	7	20
M3	4	27	7	21
M4	0	7	0	10

437

438

439 4. Discussion

440 4.1. Methodology

441 This review is novel in that it inventories an unprecedented number of ESAM from all over the world.
 442 Like Rasmussen et al. (2017) and Lampridi et al. (2019), we applied a systematic review, beginning
 443 with four search equations, whereas previously published reviews used a snowballing approach (e.g.
 444 de Olde et al., 2016). However, Rasmussen claimed their aim was not to perform a systematic review
 445 but only analyse a “robust sample from the literature”, although they did implement a search equation.
 446 Badampudi et al. (2015) compared the efficiency of the systematic and snowballing approaches and
 447 did not definitively conclude in favour of one or the other. Indeed, Wustenberghs et al. (2015)
 448 observed that due to ever-expanding methods, “any effort attempting an inventory of assessments can
 449 therefore at best be comprehensive, but not exhaustive”. To make our inventory more comprehensive,

450 we completed the inventory with two databases containing more than 100 methods, built by the
451 snowball approach (Wohlin, 2014). In both the PRISMA framework (Moher et al., 2010) and the
452 framework for systematic review for agricultural science developed by Koutsos et al. (2019),
453 additional reviews may be added manually to fill gaps (Babí Almenar et al., 2021). For our review,
454 27% of the ESAM come from manual additions. Moreover, unlike de Olde et al. (2016), we analysed
455 the whole inventory to obtain much more solid and representative results than with those previously
456 obtained. De Olde and Wustenberghs carried out four (De Olde et al., 2016) to 53 samples
457 (Wustenberghs et al. 2015).

458
459 A second novel aspect of this review lies in our aim to clearly identify and characterize conceptual and
460 methodological trends across this broad inventory and to identify gaps in those domains. Unlike de
461 Olde et al. (2016) and Marchand et al. (2014), we were not attempting to compare methods in order to
462 help end-users select an operational method. Thus, we did not include descriptors on the type of data,
463 the time needed for implementation etc. to evaluate feasibility. In all the cited reviews, we noticed a
464 mix of real descriptors (e.g. assessment purpose, scales etc.) and evaluation criteria (user-friendliness,
465 validation of results, transparency etc.; see Dale and Beyeler (2001); Binder et al. (2010); Bockstaller
466 et al. (2009); Bonisoli et al. (2018); Marchand et al. (2014)). However, in our study, some descriptors
467 or themes could also be used to evaluate and help users select a method (as we discuss in section 4.3.).
468 The Excel files provided in Tables D and F in the supplementary materials may be used as a database
469 to select methods meeting a research query, based on a set of descriptors (e.g. methods based on LCA
470 for arable crops at farm level) or themes. Our study is also much more comprehensive and identifies
471 43 descriptors. This is considerably more than the eight to 30 descriptors cited in the reviews by de
472 Olde et al. (2016); Eichler Inwood et al. (2018); Lampridi et al. (2019); Schader et al. (2014); and
473 Wustenberghs et al. (2015b). Another original aspect is the secondary focus of our review, which
474 looks at the environmental themes assessed by each method at a finer resolution than only at the three
475 sustainability dimensions (economic, environmental, social), as in de Olde et al. (2016) or
476 Wustenberghs et al. (2015b). The variety of topics or themes were added by Schader et al. (2014)
477 without any qualitative analysis. Schader et al., just like de Olde et al. (2016), also performed a theme
478 coverage analysis by comparing similarities and differences between a few methods but without
479 looking at specific themes. Eichler Inwood et al. (2018) carried out a qualitative analysis with the
480 descriptor “Efforts to identify agro-ecosystem processes or patterns” and provided general information
481 on the assessment themes, but not a detailed list of themes by methods. This part of our review is
482 inspired by Rasmussen et al. (2017), who limited their study to tropical areas and thus obtained a
483 much a smaller inventory. Rasmussen et al. worked on indicators while we tried to be more generic by
484 using themes. This led to differences as analysed in section 2.4. However, in our reviewed methods,
485 there was not always a clear difference between indicators and themes. For example, in some ESAM
486 (Röös et al., 2014; Santiago-Brown et al., 2015), indicator naming is very vague and similar to a
487 theme, although the ESAM is not structured by theme. Lastly, we were able to use the cluster analyses
488 to synthesize information by going further than a descriptor-by-descriptor analysis to obtain general
489 results. Such a step was recommended by Wustenberghs et al. (2015), who were unable to do so for
490 lack of time. Recently, Gésan-Guiziou et al. (2020) showed the potential of such an analysis by
491 applying it to the review of multi-criteria decision aid methods in agrifood systems.

492 **4.2. Results**

493 This comprehensive review demonstrates that several items that should characterize an assessment
494 method were poorly described. Indeed, several authors stressed the need to clarify “preliminary
495 choices” (Bockstaller et al., 2008) or “basic features” (Marchand et al., 2014) before any method
496 selection or design to ensure that methods will fit the end users’ needs. In particular, the purpose and
497 end users of the methods were not always specified; in this case, we assumed that the method
498 developed remained in the scientific domain and aimed to provide knowledge to researchers. If the
499 spatial scale was defined or could be easily deduced, the time scale was not always explicitly defined.
500 The calendar or cropping period (from harvest to harvest) was the implicit temporal resolution and the
501 extent depended on the project duration. Following Smyth and Dumanski (1995), we added a
502 descriptor on the expected time period for which a system is assessed to be sustainable. This was in
503 accordance with a sustainability definition class given by Hansen (1996), described as “an ability to

504 continue”. Indeed, sustainability is intrinsically linked to a temporal notion. However, only two
505 methods explicitly addressed this point (Berroterán and Zinck, 1997; Sands and Podmore, 2000). The
506 first referred directly to the scale of Smyth and Dumanski (1995) while the second explored different
507 time scales using an aggregated indicator, which they obtained by aggregating 15 predictive indicators
508 calculated with the EPIC model. We expected to see transparency issues with weighting, which is a
509 subjective exercise that should be as transparent as possible (Gan et al., 2017; Greco et al., 2019), but
510 results showed that weighting was transparent for most methods, and that method developers do take
511 the need for transparency into consideration.

512 However, the main need for clarification lies with terminology. Overcoming the inconsistent usage
513 would require an ontology (Konys, 2018), but this goes beyond the scope of this article. One important
514 point is that the difference between the conceptual item used to select an indicator (e.g. theme,
515 attribute or criterion) and the item providing the assessment (e.g. indicator, metric) is not entirely
516 clear, and sometimes even transposed. For example, for method 22, “indicator” is used for both the
517 conceptual and assessment items. For method 10, “attribute” addresses the conceptual item and
518 “indicator” the metric assessing it, while for method 23, “indicator” is related to the conceptual item
519 and “attribute” to the metric assessing it. Regardless, the conceptual framework behind the method
520 should not be limited to a list of indicators or dimensions. Three terms emerged in Figure B.7:
521 “impact”, “criterion” and “theme”. “Impact” comes from the LCA branch, and is related to the causal
522 chain for the environmental dimension. “Criterion” comes from operational research, or multi-criteria
523 decision analysis (MCDA) (Cinelli et al., 2020). But as pointed out by Fagioli et al. (2017), in
524 operational research, the term “criterion” actually refers to the indicator used in sustainability science,
525 so we would prefer the term “theme/sub-theme”, which is quite neutral compared to “issue” and less
526 theoretical than “attribute” or “variable”. “Goal” could be also used, but this imposes an orientation on
527 the conceptual framework (see next paragraph). “Goal” also speaks mainly to a policy level, such as
528 the 17 Sustainable Development Goals (Sachs, 2012). The “dimension” level can remain a commonly
529 accepted first level. Some authors used the term “principle” as a first level of very general concepts,
530 which could be placed between “dimension” and “theme” within the “dimension/(sub)themes”
531 framework.

532 With regard to the conceptual framework, most of the ESAM are based on an explicit or implicit goal-
533 oriented framework, although once again, in many cases “goals” are not explicitly defined across the
534 262 ESAM. Alternatively, Bossel (2000) argued that sustainability assessment should not depend on
535 an “intuitive” selection of indicators (or themes) and therefore proposed a list of seven basic
536 “orientors” like system existence, security, freedom etc. In fact, there is no general consensus on the
537 properties as reviewed by López-Ridaura et al. (2005), who proposed other options such as
538 “productivity” or “equity” in the MESMIS framework. In any case, this systemic property-based
539 approach is too theoretical for practical application by farmers, and the selection of indicators linked to
540 such properties is not obvious (Alkan Olsson et al., 2009). However, resilience and robustness issues
541 have become increasingly important, and farmers are taking more interest in them (see the discussion
542 below on this point). To reconcile this dual orientation, Zahm et al. (2018) linked an assessment based
543 on goals and on properties (productivity, robustness, autonomy, territorial anchorage and global
544 responsibility) in the latest version (v.4) of the IDEA method.

545 With regard to indicators, it should be noted that a large majority of the ESAM used causal indicators
546 that were mostly based on management data, even in goal-oriented ESAM (Figures 4A and 3B). For
547 example, a theme such as “water quality” is assessed using a causal indicator (e.g. “amount of
548 pesticide active ingredient”) while “preservation of biodiversity” is assessed using the same indicator
549 or a percentage of semi-natural habitat. As Bockstaller et al. (2015) and Repar et al. (2017) pointed
550 out, positioning such indicator at the beginning of the causal-effect chain raises a problem of
551 predictive quality because the actual effect assessed by the indicator is not directly related to the
552 indicator. This was clearly demonstrated for pesticide risk indicators in relation to water quality
553 (Pierlot et al., 2017). For biodiversity, some causal indicators were partially correlated to species
554 counting for a very broad range of land use types (Billeter et al., 2008) but not within one type of land
555 use such as arable farming. Such validation studies should be conducted theme by theme (see also
556 Buczko and Kuchenbuch, 2010). Nevertheless, many methods use causal indicators because they are

557 easier to implement than measuring effects or calculating a predictive indicator requiring more data.
558 Another reason could be the role of empirical knowledge in creating indicators and the adaptation of
559 these indicators by users to their situation, which was highlighted in Girard et al. (2015). In several
560 methods that combined causal and measured or predictive effect indicators, the measurement or
561 calculation remained quite easy and accessible to non-specialists (e.g. soil observation of erosion
562 marks in fields, without measuring the amount of soil loss as a reference).

563
564 Another indicator-related issue is the selection of a reference value. The majority of ESAM use
565 reference values and two-thirds of those values are absolute reference values, which is a positive point.
566 Absolute references are used to make an “absolute” assessment such as “the system is or is not
567 sustainable”. In the case of a relative reference, the system will be considered as “better or worse” than
568 a “reference system”, but this does not mean that the system is sustainable, since the outcome will
569 depend on the reference system (see Acosta-Alba and Van der Werf, (2011). Until recently, most LCA
570 studies compared an alternative option (generally a product) to a reference option and hence used a
571 relative reference. Bjørn et al. (2015) also advocated for absolute references to avoid any
572 greenwashing. Since LCA is based on quantitative impact indicators, some authors tried to adapt the
573 planetary boundaries based on available scientific knowledge (Rockstrom et al., 2009; Steffen et al.,
574 2015). However, this is not possible for causal indicators or qualitative indicators as in DEXiPM
575 (Pelzer et al., 2012). Therefore, most absolute reference values in the methods are provided by
576 “experts”, i.e. authors or a group of external experts with stakeholder participation in some cases (15%
577 of the ESAM).

578
579 Aggregation is another point that inevitably leads to debate in the scientific community (Greco et al.,
580 2019; Nardo et al., 2005; Pollesch and Dale, 2015, 2016). Nardo et al. went so far as to say there were
581 “two camps in the literature: aggregators versus non-aggregators”. Most of the methods did not
582 include an aggregation step, which may reflect careful consideration by ESAM developers; however,
583 it may also be linked to the method purpose, which in most cases was to provide knowledge, and
584 therefore to shed insight on agricultural system performance. One result that deserves attention is the
585 weak penetration of methods from operational research such as MAUT, AHP, mixed models and the
586 outranking method (Sadok et al., 2008), and that the sum of scores is the most-used aggregation
587 method, which raises concerns about the independence of indicators, compensatory rules between
588 indicators and subjectivity (Gan et al., 2017; Nardo et al., 2005). Sadok et al. (2008) called for the use
589 of mixed models that partially characterize the M3 cluster (see Table G.1). But without linking to a
590 fuzzy logic, such aggregation suffers from the threshold effect (Bockstaller et al., 2017).

591
592 The cluster analysis brought clarity to the “zoo” of SAMs (Pintér et al., 2012). The first two clusters
593 encompass methods (M1 and M2, see Table G.1) which are generally operational with causal
594 indicators as previously discussed. However, cluster M1 targets farmers predominantly through grey
595 literature publications, whereas cluster M2 is developed for researchers and published in peer-
596 reviewed papers. Both clusters cover what Repar et al. (2017) called “conventional assessment
597 methods”. M4 is characterized by LCA-based methods and confirms this specific branch, which was
598 also identified by Ness et al. (2007) and Singh et al. (2012). Cluster M3 deserves further attention as it
599 is dominated by *ex ante* assessment methods. A review by Sadok et al. (2008) showed that
600 agronomists advocate for such methods to design new field cropping systems. The MASC (Sadok et
601 al., 2009) and DEXiPM (Pelzer et al., 2012) are typical methods found in this cluster. Despite the lack
602 of a clear definition of the second cluster theme (T2), cross-checking of both cluster analyses provides
603 interpretable results. It seems that ESAM developers in cluster M1 explored all cluster themes. This
604 could be due to data availability (Repar et al. 2017) and evaluation context. The link between clusters
605 M2 and T1 shows that operational methods in many cases focus on management themes, although M2
606 shows the same with ESAM associated with the undefined cluster T2. A non-negligible share of the
607 methods in clusters M3 and M4 are associated with T4 and mainly based on environmental themes. In
608 cluster M4, which is characterized by LCA methods, most are related to T4, described by a few
609 environmental themes. Although some T4 themes such as biodiversity have been poorly covered in
610 LCA until recently (Curran et al., 2016; Gésan-Guiziou et al., 2020), the other T4 themes are typically

611 covered by LCA. For M3, some methods such as INDIGO (Bockstaller and Girardin, 2010) and EMA
612 (Lewis and Bardon, 1998) typically encompass a few environmental themes.

613

614 **4.3. Research gaps**

615 Our review also identified several gaps that should be put on the sustainability assessment research
616 agenda. One thematic gap concerns ecosystem services, which were generally neglected by the
617 methods, although publications on this theme have increased tremendously in recent years (Vihervaara
618 et al., 2010). This is likely because this concept was not sufficiently disseminated in sustainability
619 science, as was the case with the ecosystem services and the multifunctional agriculture communities
620 (Huang et al., 2015). Therond et al. (2017) highlighted the role of ecosystem services as a key factor in
621 classifying agriculture types. But it seems that there some overlap between ecosystem services and
622 “classical” environmental themes in some frameworks, such as in Tsonkova et al. (2014) or Gallardo
623 et al. (2016), who both classified the theme “erosion” as a service, while Tsonkova et al. considered
624 NO₃ content in water a service. In fact, the impacts result from an interaction of ecosystem processes
625 and anthropic factors. This is illustrated by Fegeaus et al. (2012), who proposed two indicators that
626 could be used when considering water quality as an ecosystem service: bare soil (absence of soil
627 cover) and the amount of an agricultural pollutant. The first is an ecosystem service indicator that
628 assesses the role of a process (soil cover), while the second assesses the anthropic driver and should be
629 linked to the impact category. With regard to climate change, GHG emissions are assessed, but only a
630 few cases consider the ecosystem service of carbon storage, such as CAP2’ER (Moreau, 2016), and
631 even fewer consider albedo (soil reflectance) as pointed out by Carrer et al. (2018). Thus, we suggest
632 clearly integrating ecosystem services into SAMs and separately from impacts. This will help provide
633 insights on system functioning and the role of the ecosystems in mitigating or amplifying an impact, as
634 well as offer more complete information on the negative impacts of agriculture along with positive
635 contributions in relation to its multifunctionality (Huang et al., 2015).

636 Another thematic gap deals with pollution from dissolved organic carbon and nitrogen due to their
637 impact on water quality (Stanley et al., 2012). However, data are still scarce. The studied methods did
638 not deal with emerging themes such as pollution from antibiotics and other medicines use in
639 agriculture (Du and Liu, 2012) or microplastics (Horton et al., 2017), and the interactions between
640 both types of pollutants received even less attention (Lu et al., 2020). With regard to antibiotics, Lynch
641 et al. (2019) stated in their research of sustainability indicators for Ireland that the only indicators that
642 could be proposed would be based on national statistics of antibiotic use. This would be a causal
643 indicator which provides no information on the risk of antibiotic resistance. Agricultural soils are not
644 only major receptors of microplastics through sewage sludge but also a source of microplastics
645 through plastic mulches and polytunnels. Data exist for sewage sludge but not for mulches and
646 polytunnels (Horton et al., 2017). The impacts of microplastics could be again assessed by causal
647 indicators such as the amount of plastic used or amount of waste. In our review, 12% of the methods
648 covered waste management and some addressed plastic use (e.g. FAO, 2013; Melo and Wolf, 2005;
649 Schader et al., 2014), while others considered it to be of low importance and excluded it (Kuneman
650 and Fellus, 2014). Vanham et al. (2019) proposed a measured indicator of plastic waste density in
651 coastal water. But no authors referred directly to the impact of microplastics.

652 Another major gap deals with interaction between indicators. This subject was very poorly covered by
653 the methods (just four of 262), although Binder et al. (2010) recommended it as a criterion to evaluate
654 the systemic dimension and de Olde et al. (2016) used it to compare methods. Furthermore, Gan et al.
655 (2017) noted the need to label interactions between indicator results before choosing an aggregation
656 method since additive aggregation methods are not suitable when there are interactions. In our review,
657 interactions between indicators were not addressed at this level aside from studying relationships
658 between indicators, namely causal indicators used to assess management and effect indicators (Smith
659 et al., 2000; Trivino-Tarradas et al., 2019). Binder et al. (2012) proposed plotting the range of
660 variation in indicator pairs to identify the sustainability solution space, i.e. the space of the plot for
661 which indicators are considered as sustainable based on the reference value. Finally, Küstermann et al.
662 (2010) proposed a systemic representation of different indicators assessing nitrogen flows and losses
663 in farming systems, but this was a unique case for nitrogen only.

664 One final methodological gap pertains to scales, known to be “a difficult issue” (Scholes et al., 2013).
665 We observed that for the spatial dimension, most of the ESAM calculated indicators at a given
666 resolution (farm level for those that were most represented) and extent (e.g. territory for those most
667 represented) as shown in Figure 5. But once again, very few explicitly detailed an upscaling or
668 downscaling procedure (four of 262 methods). While Tzilivakis and Lewis (2004) proposed a
669 disaggregation method to make the indicators at country level more meaningful at farm level, other
670 authors proposed an aggregation method. Delbaere and Serradilla (2004) derived statistics from
671 indicators calculated at a lower level to obtain a higher-level indicator expressed as a percentage of
672 area associated with a value class of the lower-level indicator (e.g. for the erosion indicator calculated
673 for a 50 km x 50 km grid cell, for the country level they obtained an erosion class covering 50% of the
674 country area). This is a statistical upscaling while Bartolini et al. (2007) carried out a linear
675 aggregation by weighted sum of indicator results of farms according to the share of usable agricultural
676 area belonging to that farm type. Chopin et al. (2017) calculated statistics just as Delbaere and
677 Serradilla (2004) (e.g. for pesticide risk indicators) or by using a sum of the indicator values (e.g. farm
678 income) at regional level. It is noticeable that non-linear aggregation as proposed by Dalgaard et al.
679 (2003) or Scholes et al. (2013) was not implemented. This can be explained by the need for more
680 knowledge on processes, although Dalgaard et al. (2003) suggested an easy example for energy
681 consumption, taking into account the distance from farm to field rather than just field size. Lastly, only
682 one method (Smith et al., 2000) integrated spatially explicit information in the calculation of
683 indicators; this was also done for hydrological aspects by Wohlfahrt et al. (2010) using the INDIGO
684 method (Bockstaller and Girardin, 2010). However, this requires additional data collection and
685 processing efforts that exceed the skills and means of most ESAM developers.

686 With regard to the temporal dimension, we noted that the calendar or cropping period (from harvest to
687 harvest) is often the implicit temporal resolution. Two studies explicitly defined a long-term scale
688 (Berroterán and Zinck, 1997; Sands and Podmore, 2000). Berroterán and Zinck (1997) carried out an
689 *ex post* assessment at national scale (Venezuela) while Sands and Podmore (2000) opted for an *ex ante*
690 assessment by modelling. However, the authors did not integrate dynamic concepts such as resilience
691 or robustness, which can be linked to sustainability, into their assessments. In our review, eight
692 methods addressed robustness/resilience, including Zahm (2008) in the IDEA v.4 method, and three
693 included adaptability/flexibility (see López-Ridaura et al. (2005)). Although the authors of these eight
694 methods defined robustness and adaptability in a dynamic way, they only proposed a list of static
695 indicators (calculated for one year). In fact, they followed one of the approaches reviewed by Urruty et
696 al. (2016) which is based on a set of indicators describing internal system features. Urruty et al. (2016)
697 identified two other approaches: one based on an analysis of output variability and the other on an
698 analysis of system behaviour in the event of a perturbation or shock. In a recent review, Dardonville et
699 al. (2020) identified about 50 papers that quantitatively assessed resilience/robustness/vulnerability.
700 However, all of those papers only dealt with one or two sustainability themes (in many cases
701 profitability or productivity). There is a need to continue such studies, but they need to take a multi-
702 criteria approach to a set of sustainability indicators. Starting with the study by Sands and Podmore
703 (2000), one possibility would be to integrate one or several years with a strong perturbation (e.g.
704 severe drought or water excess) in climatic series and model the system behaviour to calculate the time
705 of system recovery, which could be an indicator of resilience. Such a study would require a modelling
706 effort and adapted datasets.

707

708 **5. Conclusion**

709 In this study, we performed a review of unprecedented size of environmental sustainability assessment
710 methods from all over the world. The main aim was not to look at comparison criteria, even if the
711 databases describing the methods may make it possible. Rather, we sought to thoroughly analyse the
712 conceptual and methodological framework behind these methods as well as the environmental themes
713 they covered. We identified four clusters of method descriptors and four clusters of environmental
714 themes. Two of the method clusters – one characterized by *ex ante* methods to select, classify or rank
715 at field level and one characterized by LCA – mainly deal with limited themes covering the main
716 environmental issues (climate change, water quality, soil erosion etc.).

717 ESAM developers can take away several general recommendations from our review. The first is a
718 need for clear, consistent terminology. We showed that even the most frequently used term “indicator”
719 does not have a consistent definition. Furthermore, ESAM should have explicitly stated purposes,
720 target users and temporal scales. We observed many studies where this information was only implicit,
721 although authors did make an effort to be transparent when they associated a reference value or a
722 weight with specific indicators.

723 Additional efforts must be made in selecting indicators positioned on the causal chain at emission
724 level or for impacts, such as in LCA. An alternative would be to select causal indicators based on
725 process knowledge. Additional research gaps include considering indicator interactions before
726 aggregation when necessary, going beyond a simple sum of scores when aggregating indicators, and
727 dealing with upscaling and spatialization issues. A further challenge that remains is to integrate a
728 dynamic approach based on the study of resilience or robustness, not only for a single criterion such as
729 productivity or profitability but with multiple criteria. In the future, ESAM must also address
730 emerging environmental issues such as pollution from antibiotics or microplastics and the integration
731 of ecosystem services. Lastly, a similar study should be conducted for the social and economic
732 dimensions with some adaptations, especially with regard to indicator typology in relation to the
733 causal chain and themes.

734 **References**

- 735 Acosta-Alba, I., Van der Werf, H., 2011. The Use of Reference Values in Indicator-Based Methods for
736 the Environmental Assessment of Agricultural Systems. *Sustainability* 3, 424–442.
737 <https://doi.org/10.3390/su3020424>
- 738 Alkan Olsson, J., Bockstaller, C., Stapleton, L., Knapen, R., Therond, O., Turpin, N., Geniaux, G.,
739 Bellon, S., Pinto Correia, T., Bezlepkina, I., Taverne, M., Ewert, F., 2009. Indicator frameworks
740 supporting ex-ante impact assessment of new policies for rural systems; a critical review of a
741 goal oriented framework and its indicators. *Environ. Sci. Policy* 12, 562–572.
- 742 Alkan Olsson, Johanna, Bockstaller, C., Stapleton, L.M., Ewert, F., Knapen, R., Therond, O., Geniaux,
743 G., Bellon, S., Correia, T.P., Turpin, N., Bezlepkina, I., 2009. A goal oriented indicator
744 framework to support integrated assessment of new policies for agri-environmental systems.
745 *Environ. Sci. Policy* 12, 562–572. <https://doi.org/10.1016/j.envsci.2009.01.012>
- 746 Babí Almenar, J., Elliot, T., Rugani, B., Philippe, B., Navarrete Gutierrez, T., Sonnemann, G.,
747 Geneletti, D., 2021. Nexus between nature-based solutions, ecosystem services and urban
748 challenges. *Land use policy* 100, 104898. <https://doi.org/10.1016/j.landusepol.2020.104898>
- 749 Badampudi, D., Wohlin, C., Petersen, K., 2015. Experiences from using snowballing and database
750 searches in systematic literature studies. *ACM Int. Conf. Proceeding Ser.* 27-29-April.
751 <https://doi.org/10.1145/2745802.2745818>
- 752 Bartolini, F., Bazzani, G.M., Gallerani, V., Raggi, M., Viaggi, D., 2007. The impact of water and
753 agriculture policy scenarios on irrigated farming systems in Italy: An analysis based on farm
754 level multi-attribute linear programming models. *Agric. Syst.* 93, 90–114.
755 <https://doi.org/10.1016/j.agsy.2006.04.006>
- 756 Bell, S., Morse, S., 2001. Breaking through the glass ceiling: Who really cares about Sustainability
757 Indicators? *Local Environ.* 6, 291–309. <https://doi.org/DOI: 10.1080/13549830120073284>
- 758 Berroterán, J.L., Zinck, J.A., 1997. Indicators of Agricultural Sustainability at the National Level: A
759 Case Study of Venezuela.
- 760 Billeter, R., Liira, J., Bailey, D., Bugter, R., Arens, P., Augenstein, I., Aviron, S., Baudry, J., Bukacek,
761 R., Burel, F., Cerny, M., De Blust, G., De Cock, R., Diekötter, T., Dietz, H., Dirksen, J.,
762 Dormann, C., Durka, W., Frenzel, M., Hamersky, R., Hendrickx, F., Herzog, F., Klotz, S.,
763 Koolstra, B., Lausch, A., Le Coeur, D., Maelfait, J.P., Opdam, P., Roubalova, M., Schermann,
764 A., Schermann, N., Schmidt, T., Schweiger, O., Smulders, M.J.M., Speelmans, M., Simova, P.,

- 765 Verboom, J., Van Wingerden, W.K.R.E., Zobel, M., Edwards, P.J., 2008. Indicators for
766 biodiversity in agricultural landscapes: A pan-European study. *J. Appl. Ecol.* 45, 141–150.
767 <https://doi.org/10.1111/j.1365-2664.2007.01393.x>
- 768 Binder, C.R., Feola, G., Steinberger, J.K., 2010. Considering the normative, systemic and procedural
769 dimensions in indicator-based sustainability assessments in agriculture. *Environ. Impact Assess.*
770 *Rev.* 30, 71–81. <https://doi.org/10.1016/j.eiar.2009.06.002>
- 771 Binder, C.R., Schmid, A., Steinberger, J.K., 2012. Sustainability solution space of the Swiss milk
772 value added chain. *Ecol. Econ.* 83, 210–220. <https://doi.org/10.1016/j.ecolecon.2012.06.022>
- 773 Bjørn, A., Diamond, M., Owsianiak, M., Verzat, B., Hauschild, M.Z., 2015. Strengthening the link
774 between life cycle assessment and indicators for absolute sustainability to support development
775 within planetary boundaries. *Environ. Sci. Technol.* 49, 6370–6371.
776 <https://doi.org/10.1021/acs.est.5b02106>
- 777 Bockstaller, C., Beauchet, S., Manneville, V., Amiaud, B., Botreau, R., 2017. A tool to design fuzzy
778 decision trees for sustainability assessment. *Environ. Model. Softw.* 97, 130–144.
779 <https://doi.org/10.1016/j.envsoft.2017.07.011>
- 780 Bockstaller, C., Feschet, P., Angevin, F., 2015. Issues in evaluating sustainability of farming systems
781 with indicators. *OCL - Oilseeds fats* 22. <https://doi.org/10.1051/ocl/2014052>
- 782 Bockstaller, C., Girardin, P., 2010. MODE DE CALCUL DES INDICATEURS AGRICULTURELS
783 ENVIRONNEMENTAUX DE LA METHODE INDIGO ®.
- 784 Bockstaller, C., Guichard, L., Keichinger, O., Girardin, P., Galan, M.B., Gaillard, G., 2009.
785 Comparison of methods to assess the sustainability of agricultural systems. A review. *Agron.*
786 *Sustain. Dev.* 29, 223–235. <https://doi.org/doi:10.1051/agro:2008058>
- 787 Bockstaller, Christian, Guichard, L., Makowski, D., Aveline, A., Girardin, P., Plantureux, S., 2009.
788 Agri-environmental indicators to assess cropping and farming systems: A review. *Sustain. Agric.*
789 28, 725–738. https://doi.org/10.1007/978-90-481-2666-8_44
- 790 Bockstaller, C., Guichard, L., Makowski, D., Aveline, A., Girardin, P., Plantureux, S., 2008. Agri-
791 environmental indicators to assess cropping and farming systems. A review. *Agron. Sustain.*
792 *Dev.* 28, 139–149. <https://doi.org/doi:10.1051/agro:2007052>
- 793 Bonisoli, L., Galdeano-Gómez, E., Piedra-Muñoz, L., 2018. Deconstructing criteria and assessment
794 tools to build agri-sustainability indicators and support farmers' decision-making process. *J.*
795 *Clean. Prod.* 182, 1080–1094. <https://doi.org/10.1016/j.jclepro.2018.02.055>
- 796 Bossel, H., 2000. Policy assessment and simulation of actor orientation for sustainable development.
797 *Ecol. Econ.* 35, 337–355. [https://doi.org/10.1016/S0921-8009\(00\)00218-4](https://doi.org/10.1016/S0921-8009(00)00218-4)
- 798 Buckwell, A., Uhre, A.N., Annabelle, W., Jana, P., Blum, W., Schiefer, J., Lair, G.J., Heissenhuber,
799 A., Schießl, P., Krämer, C., Haber, W., 2014. Sustainable Intensification of European
800 Agriculture A review sponsored by the RISE Foundation The. *RISE Found.* 98.
801 <https://doi.org/10.1111/j.1477-8947.1997.tb00699.x>
- 802 Buczko, U., Kuchenbuch, R.O., 2010. Environmental Indicators to Assess the Risk of Diffuse
803 Nitrogen Losses from Agriculture. *Environ. Manage.* 45, 1201–1222.
804 <https://doi.org/10.1007/s00267-010-9448-8>
- 805 Campbell, B.M., Beare, D.J., Bennett, E.M., Hall-Spencer, J.M., Ingram, J.S.I., Jaramillo, F., Ortiz, R.,
806 Ramankutty, N., Sayer, J.A., Shindell, D., 2017. Agriculture production as a major driver of the
807 earth system exceeding planetary boundaries. *Ecol. Soc.* 22. <https://doi.org/10.5751/ES-09595-220408>
- 809 Candelaria-Martínez, B., Ruiz-Rosado, O., Pérez-Hernández, P., Gallardo-López, F., Vargas-Villamil,

- 810 L., Martínez-Becerra, Á., Flota-Bañuelos, C., 2014. Sustentabilidad de los agroecosistemas de la
811 microcuenca Paso de Ovejas 1, Veracruz, México. *Cuad. Desarro. Rural* 11, 87–104.
812 <https://doi.org/10.11144/Javeriana.CDR11-73.sdsm>
- 813 Carrer, D., Pique, G., Ferlicoq, M., Ceamanos, X., Ceschia, E., 2018. What is the potential of cropland
814 albedo management in the fight against global warming? A case study based on the use of cover
815 crops. *Environ. Res. Lett.* 13. <https://doi.org/10.1088/1748-9326/aab650>
- 816 Castoldi, N., Bechini, L., 2010. Integrated sustainability assessment of cropping systems with agro-
817 ecological and economic indicators in northern Italy. *Eur. J. Agron.* 32, 59–72.
818 <https://doi.org/10.1016/j.eja.2009.02.003>
- 819 Cinelli, M., Spada, M., Kim, W., Zhang, Y., Burgherr, P., 2020. MCDA Index Tool: an interactive
820 software to develop indices and rankings. *Environ. Syst. Decis.* <https://doi.org/10.1007/s10669-020-09784-x>
- 822 Coteur, I., Marchand, F., Debruyne, L., Lauwers, L., 2019. Structuring the myriad of sustainability
823 assessments in agri-food systems: A case in Flanders. *J. Clean. Prod.* 209, 472–480.
824 <https://doi.org/10.1016/J.JCLEPRO.2018.10.066>
- 825 Curran, M., De Souza, D.M., Antón, A., Teixeira, R.F.M., Michelsen, O., Vidal-Legaz, B., Sala, S.,
826 Milà I Canals, L., 2016. How Well Does LCA Model Land Use Impacts on Biodiversity? - A
827 Comparison with Approaches from Ecology and Conservation. *Environ. Sci. Technol.*
828 <https://doi.org/10.1021/acs.est.5b04681>
- 829 Dale, V.H., Beyeler, S.C., 2001. Challenges in the development and use of ecological indicators. *Ecol.*
830 *Indic.* 1, 3–10. [https://doi.org/10.1016/S1470-160X\(01\)00003-6](https://doi.org/10.1016/S1470-160X(01)00003-6)
- 831 Dalgaard, T., Hutchings, N.J., Porter, J.R., 2003. Agroecology, scaling and interdisciplinarity. *Agric.*
832 *Ecosyst. Environ.* 100, 39–51.
- 833 Dardonville, M., Bockstaller, C., Therond, O., 2020. Review of quantitative evaluations of the
834 resilience, vulnerability, robustness and adaptive capacity of temperate agricultural systems. *J.*
835 *Clean. Prod.* 125456. <https://doi.org/10.1016/j.jclepro.2020.125456>
- 836 de Olde, E.M., Oudshoorn, F.W., Bokkers, E.A.M., Stubsgaard, A., Sørensen, C.A.G., de Boer, I.J.M.,
837 2016. Assessing the sustainability performance of organic farms in Denmark. *Sustain.* 8.
838 <https://doi.org/10.3390/su8090957>
- 839 De Olde, E.M., Oudshoorn, F.W., Sørensen, C.A.G., Bokkers, E.A.M., De Boer, I.J.M., 2016.
840 Assessing sustainability at farm-level: Lessons learned from a comparison of tools in practice.
841 *Ecol. Indic.* 66, 391–404. <https://doi.org/10.1016/j.ecolind.2016.01.047>
- 842 Delbaere, B., Serradilla, a. N., 2004. Environmental risks from agriculture in Europe. *Interpret. A J.*
843 *Bible Theol.* 184.
- 844 Du, L., Liu, W., 2012. Occurrence, fate, and ecotoxicity of antibiotics in agro-ecosystems. A review.
845 *Agron. Sustain. Dev.* 32, 309–327. <https://doi.org/10.1007/s13593-011-0062-9>
- 846 Duncombe-Wall, D., 1999. Agricultural sustainability indicators for regions of South Australia
847 Cardiac ARIA View project Investigation of Local Area Factors Related to BreastScreen New
848 South Wales View project. <https://doi.org/10.13140/RG.2.1.2887.5287>
- 849 Egea, P., Pérez y Pérez, L., 2016. Sustainability and multifunctionality of protected designations of
850 origin of olive oil in Spain. *Land use policy* 58, 264–275.
851 <https://doi.org/10.1016/j.landusepol.2016.07.017>
- 852 Eichler Inwood, S.E., López-Ridaura, S., Kline, K.L., Gérard, B., Monsalve, A.G., Govaerts, B., Dale,
853 V.H., 2018. Assessing sustainability in agricultural landscapes: A review of approaches1,2.
854 *Environ. Rev.* 26, 299–315. <https://doi.org/10.1139/er-2017-0058>

- 855 El Chami, D., Daccache, A., 2015. Assessing sustainability of winter wheat production under climate
856 change scenarios in a humid climate - An integrated modelling framework. *Agric. Syst.* 140, 19–
857 25. <https://doi.org/10.1016/j.agsy.2015.08.008>
- 858 Fagioli, F.F., Rocchi, L., Paolotti, L., Słowiński, R., Boggia, A., 2017. From the farm to the agri-food
859 system: A multiple criteria framework to evaluate extended multi-functional value. *Ecol. Indic.*
860 79, 91–102. <https://doi.org/10.1016/j.ecolind.2017.04.009>
- 861 FAO, 2013. Sustainability Assessment of Food and Agriculture systems.
- 862 Fegraus, E.H., Zaslavsky, I., Whitenack, T., Dempewolf, J., Ahumada, J.A., Lin, K., Andelman, S.J.,
863 2012. Interdisciplinary decision support dashboard: A new framework for a tanzanian
864 agricultural and ecosystem service monitoring system pilot. *IEEE J. Sel. Top. Appl. Earth Obs.*
865 *Remote Sens.* 5, 1700–1708. <https://doi.org/10.1109/JSTARS.2012.2204864>
- 866 Feola, G., Rahn, E., Binder, C.R., 2011. Suitability of pesticide risk indicators for Less Developed
867 Countries: A comparison. *Agric. Ecosyst. Environ.* 142, 238–245.
868 <https://doi.org/10.1016/j.agee.2011.05.014>
- 869 Fernandes, L.A. de O., Woodhouse, P.J., 2008. Family farm sustainability in southern Brazil: An
870 application of agri-environmental indicators. *Ecol. Econ.* 66, 243–257.
871 <https://doi.org/10.1016/j.ecolecon.2008.01.027>
- 872 Gallardo, A.L., Duarte, C., Dibo, A.P., 2016. STRATEGIC ENVIRONMENTAL ASSESSMENT
873 FOR PLANNING SUGARCANE EXPANSION: A FRAMEWORK PROPOSAL. *Ambient.*
874 *Soc.* 19, 67–92.
- 875 Gan, X., Fernandez, I.C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., Wu, J., 2017. When to use what:
876 Methods for weighting and aggregating sustainability indicators. *Ecol. Indic.* 81, 491–502.
877 <https://doi.org/10.1016/j.ecolind.2017.05.068>
- 878 Gésan-Guiziu, G., Alaphilippe, A., Aubin, J., Bockstaller, C., Boutrou, R., Buche, P., Collet, C.,
879 Girard, A., Martinet, V., Membré, J.M., Sabbadin, R., Thiollet-Scholtus, M., van der Werf,
880 H.M.G., 2020. Diversity and potentiality of multi-criteria decision analysis methods for agri-food
881 research. *Agron. Sustain. Dev.* 40, 1–11. <https://doi.org/10.1007/s13593-020-00650-3>
- 882 Girard, N., Magda, D., Astruc, J.M., Couix, N., Gross, H., P Guyon, J., Labatut, J., Poinot, Y.,
883 Saldaqui, F., 2015. Analyzing indicators for combining natural resources management and
884 production-oriented activities. *Environ. Dev. Sustain.* 17, 155–172.
885 <https://doi.org/10.1007/s10668-014-9544-7>
- 886 GIS GC HP2E, 2018. INDIC® : Une base de données sur les méthodes d'évaluation et indicateurs de
887 durabilité téléchargeable sur le site du GIS GC HP2E [WWW Document]. URL
888 [https://www.gchp2e.fr/Actualites/INDIC-R-Une-base-de-donnees-telechargeable-sur-le-site-du-](https://www.gchp2e.fr/Actualites/INDIC-R-Une-base-de-donnees-telechargeable-sur-le-site-du-GIS-GC-HP2E)
889 [GIS-GC-HP2E](https://www.gchp2e.fr/Actualites/INDIC-R-Une-base-de-donnees-telechargeable-sur-le-site-du-GIS-GC-HP2E)
- 890 Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J.,
891 Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food security: The challenge of feeding 9
892 billion people. *Science* (80-). <https://doi.org/10.1126/science.1185383>
- 893 Gomez-Limon, J.A., Riesgo, L., Gómez-Limón, J.A., 2010. Sustainability assessment of olive grove in
894 Andalusia: A methodological proposal Sustainability assessment of olive grove in Andalusia: A
895 methodological proposal 1, in: *External Cost of Farming Activities: Economic Evaluation,*
896 *Environmental Repercussions and Regulatory Framework.*
- 897 Greco, S., Ishizaka, A., Tasiou, M., Torrisi, G., 2019. On the Methodological Framework of
898 Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. *Soc.*
899 *Indic. Res.* 141, 61–94. <https://doi.org/10.1007/s11205-017-1832-9>
- 900 Haas, G., Wetterich, F., Geier, U., 2000. LCA Methodology Framework in Agriculture on the Farm

- 901 Level 1 LCA Methodology Framework in Agriculture on the Farm Level. *Korea Int. J. LCA* 5,
902 2000. <https://doi.org/10.1065/lca2000.011.038>
- 903 Hanegraaf, M.C., Biewinga, E.E., Van Der Bijl, G., 1998. ASSESSING THE ECOLOGICAL AND
904 ECONOMIC SUSTAINABILITY OF ENERGY CROPS.
- 905 Hansen, J.W., 1996. Is agricultural sustainability a useful concept? *Agric. Syst.* 50, 117–143.
906 [https://doi.org/10.1016/0308-521X\(95\)00011-S](https://doi.org/10.1016/0308-521X(95)00011-S)
- 907 Horton, A.A., Walton, A., Spurgeon, D.J., Lahive, E., Svendsen, C., 2017. Microplastics in freshwater
908 and terrestrial environments: Evaluating the current understanding to identify the knowledge
909 gaps and future research priorities. *Sci. Total Environ.* 586, 127–141.
910 <https://doi.org/10.1016/j.scitotenv.2017.01.190>
- 911 Huang, J., Tichit, M., Poulot, M., Darly, S., Li, S., Petit, C., Aubry, C., 2015. Comparative review of
912 multifunctionality and ecosystem services in sustainable agriculture. *J. Environ. Manage.* 149,
913 138–147. <https://doi.org/10.1016/j.jenvman.2014.10.020>
- 914 Husson, F., Lê, S., Pages, J., 2010. Exploratory Multivariate Analysis by Example Using R - Francois
915 Husson, Sebastien Le, Jérôme Pagès - Google Livres [WWW Document]. URL
916 [https://books.google.fr/books?hl=fr&lr=&id=nLrODgAAQBAJ&oi=fnd&pg=PP1&dq=Husson,
917 +F.,+Lê,+S.,+Pages,+J.,+2010.+Exploratory+multivariate+analysis+by+example+using+R,+vol
918 ume+20105550+of+Chapman+%26+Hall/CRC+Computer+Science+%26+Data+Analysis.+CRC
919 +Press+30,+1+\(accessed+12.4.20\)](https://books.google.fr/books?hl=fr&lr=&id=nLrODgAAQBAJ&oi=fnd&pg=PP1&dq=Husson,+F.,+Lê,+S.,+Pages,+J.,+2010.+Exploratory+multivariate+analysis+by+example+using+R,+volume+20105550+of+Chapman+%26+Hall/CRC+Computer+Science+%26+Data+Analysis.+CRC+Press+30,+1+(accessed+12.4.20)).
- 920 IPCC, 2003. Revision of the “Revised 1996 IPCC Guidelines for National Greenhouse Gas
921 Inventories, Intergovernmental Panel On Climate Change national greenhouse gas inventories
922 programme.
- 923 Konys, A., 2018. An ontology-based knowledge modelling for a sustainability assessment domain.
924 *Sustain.* 10. <https://doi.org/10.3390/su10020300>
- 925 Koutsos, T.M., Menexes, G.C., Dordas, C.A., 2019. An efficient framework for conducting systematic
926 literature reviews in agricultural sciences. *Sci. Total Environ.*
927 <https://doi.org/10.1016/j.scitotenv.2019.04.354>
- 928 Kuneman, G., Fellus, E., 2014. Towards Consistent Measurement of Sustainability at Farm level.
- 929 Küstermann, B., Christen, O., Hülsbergen, K.J., 2010. Modelling nitrogen cycles of farming systems
930 as basis of site- and farm-specific nitrogen management. *Agric. Ecosyst. Environ.* 135, 70–80.
931 <https://doi.org/10.1016/j.agee.2009.08.014>
- 932 Lampridi, M.G., Sørensen, C.G., Bochtis, D., 2019. Agricultural sustainability: A review of concepts
933 and methods. *Sustain.* <https://doi.org/10.3390/su11185120>
- 934 Laurent, C., Hulin, S., Agabriel, C., Chassaing, C., Botreau, R., Monteils, V., 2017. Co-construction of
935 an assessment method of the environmental sustainability for cattle farms involved in a Protected
936 Designation of Origin (PDO) cheese value chain, Cantal PDO. *Ecol. Indic.* 76, 357–365.
937 <https://doi.org/10.1016/j.ecolind.2017.01.035>
- 938 Lebacqz, T., Baret, P. V., Stilmant, D., 2013. Sustainability indicators for livestock farming. A review.
939 *Agron. Sustain. Dev.* 33, 311–327. <https://doi.org/10.1007/s13593-012-0121-x>
- 940 Levitan, L., 2000. “How to” and “why”: assessing the enviro-social impacts of pesticides. *Crop*
941 *Prot.* 19, 629–636.
- 942 Lewis, K.A., Bardon, K.S., 1998. A computer-based informal environmental management system for
943 agriculture Tel: 44 (0) 1707 284582 Fax: 44 (0) 1707 285258 Email: K.A.Lewis@herts.ac.uk 2
944 A computer-based informal environmental management system for agriculture.
- 945 López-Ridaura, S., Van Keulen, H., Van Ittersum, M.K., Leffelaar, P.A., 2005. Multiscale

- 946 methodological framework to derive criteria and indicators for sustainability evaluation of
 947 peasant natural resource management systems. *Environ. Dev. Sustain.* 7, 51–69.
 948 <https://doi.org/10.1007/s10668-003-6976-x>
- 949 Lu, X.M., Lu, P.Z., Liu, X.P., 2020. Fate and abundance of antibiotic resistance genes on
 950 microplastics in facility vegetable soil. *Sci. Total Environ.* 709, 136276.
 951 <https://doi.org/10.1016/j.scitotenv.2019.136276>
- 952 Lynch, J., Donnellan, T., Finn, J.A., Dillon, E., Ryan, M., 2019. Potential development of Irish
 953 agricultural sustainability indicators for current and future policy evaluation needs. *J. Environ.*
 954 *Manage.* <https://doi.org/10.1016/j.jenvman.2018.09.070>
- 955 Marchand, F., Debruyne, L., Triste, L., Gerrard, C., Padel, S., Lauwers, L., 2014. Key characteristics
 956 for tool choice in indicator-based sustainability assessment at farm level. *Ecol. Soc.* 19.
 957 <https://doi.org/10.5751/ES-06876-190346>
- 958 Melo, C.J., Wolf, S.A., 2005. Empirical assessment of eco-certification: The case of Ecuadorian
 959 Bananas. *Organ. Environ.* 18, 287–317. <https://doi.org/10.1177/1086026605279461>
- 960 Mitchell, G., May, A., Mc Donald, A., 1995. PICABUE: a methodological framework for the
 961 development of indicators of sustainable development. *Int. J. Sustain. Dev. World Ecol.* 2, 104–
 962 123.
- 963 Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2010. Preferred reporting items for systematic
 964 reviews and meta-analyses: The PRISMA statement. *Int. J. Surg.* 8, 336–341.
 965 <https://doi.org/10.1016/j.ijsu.2010.02.007>
- 966 Moraine, M., Duru, M., Therond, O., 2017. A social-ecological framework for analyzing and
 967 designing integrated crop-livestock systems from farm to territory levels. *Renew. Agric. Food*
 968 *Syst.* 32, 43–56. <https://doi.org/10.1017/S1742170515000526>
- 969 Moreau, 2016. Guide methologique CAP2ER.
- 970 Nardo, M., Saisana, M., Tarantola, A., Stefano, S., 2005. Tools for Composite Indicators Building 1–
 971 134.
- 972 Nasca, J.A., Toranzos, M., Banegas, N.R., 2006. Evaluación de la sostenibilidad de dos modelos
 973 ganaderos de la llanura deprimida salina de Tucumán, Argentina.
- 974 Ness, B., Urbel-Piirsalu, E., Anderberg, S., Olsson, L., 2007. Categorising tools for sustainability
 975 assessment. *Ecol. Econ.* 60, 498–508. <https://doi.org/10.1016/j.ecolecon.2006.07.023>
- 976 Niemeijer, D., de Groot, R.S., 2008. A conceptual framework for selecting environmental indicator
 977 sets. *Ecol. Indic.* 8, 14–25. <https://doi.org/10.1016/j.ecolind.2006.11.012>
- 978 Payraudeau, S., Van Der Werf, H.M.G., 2005. Environmental impact assessment for a farming region:
 979 A review of methods. *Agric. Ecosyst. Environ.* 107, 1–19.
 980 <https://doi.org/10.1016/j.agee.2004.12.012>
- 981 Pelzer, E., Fortino, G., Bockstaller, C., Angevin, F., Lamine, C., Moonen, C., Vasileiadis, V., Guérin,
 982 D., Guichard, L., Reau, R., Messéan, A., 2012. Assessing innovative cropping systems with
 983 DEXiPM, a qualitative multi-criteria assessment tool derived from DEXi. *Ecol. Indic.* 18, 171–
 984 182. <https://doi.org/10.1016/j.ecolind.2011.11.019>
- 985 Pierlot, F., Marks-Perreau, J., Réal, B., Carluer, N., Constant, T., Lioeddine, A., van Dijk, P., Villerd,
 986 J., Keichinger, O., Cherrier, R., Bockstaller, C., 2017. Predictive quality of 26 pesticide risk
 987 indicators and one flow model: A multisite assessment for water contamination. *Sci. Total*
 988 *Environ.* 605–606, 655–665. <https://doi.org/10.1016/j.scitotenv.2017.06.112>
- 989 Pintér, L., Hardi, P., Martinuzzi, A., Hall, J., 2012. Bellagio STAMP: Principles for sustainability
 990 assessment and measurement. *Ecol. Indic.* 17, 20–28.

- 991 <https://doi.org/10.1016/j.ecolind.2011.07.001>
- 992 Pollesch, N., Dale, V.H., 2015. Applications of aggregation theory to sustainability assessment. *Ecol.*
993 *Econ.* 114, 117–127. <https://doi.org/10.1016/j.ecolecon.2015.03.011>
- 994 Pollesch, N.L., Dale, V.H., 2016. Normalization in sustainability assessment: Methods and
995 implications. *Ecol. Econ.* 130, 195–208. <https://doi.org/10.1016/j.ecolecon.2016.06.018>
- 996 Ponisio, L.C., M’Gonigle, L.K., Mace, K.C., Palomino, J., de Valpine, P., Kremen, C., 2015.
997 Diversification practices reduce organic to conventional yield gap. *Proc. R. Soc. B-Biological*
998 *Sci.* 282. <https://doi.org/10.1098/rspb.2014.1396>
- 999 Pope, J., Annandale, D., Morrison-Saunders, A., 2004. Conceptualising sustainability assessment.
1000 *Environ. Impact Assess. Rev.* 24, 595–616. <https://doi.org/10.1016/j.eiar.2004.03.001>
- 1001 Purtauf, T., Thies, C., Ekschmitt, K., Wolters, V., Dauber, J., 2005. Scaling properties of multivariate
1002 landscape structure. *Ecol. Indic.* 5, 295–304.
- 1003 Rasmussen, L.V., Bierbaum, R., Oldekop, J.A., Agrawal, A., 2017. Bridging the practitioner-
1004 researcher divide: Indicators to track environmental, economic, and sociocultural sustainability
1005 of agricultural commodity production. *Glob. Environ. Chang.* 42, 33–46.
1006 <https://doi.org/10.1016/j.gloenvcha.2016.12.001>
- 1007 Repar, N., Jan, P., Dux, D., Nemecek, T., Doluschitz, R., 2017. Implementing farm-level
1008 environmental sustainability in environmental performance indicators: A combined global-local
1009 approach. *J. Clean. Prod.* 140, 692–704. <https://doi.org/10.1016/j.jclepro.2016.07.022>
- 1010 Riley, J., 2001. The indicator explosion: Local needs and international challenges. *Agric. Ecosyst.*
1011 *Environ.* 87, 119–120. [https://doi.org/10.1016/S0167-8809\(01\)00271-7](https://doi.org/10.1016/S0167-8809(01)00271-7)
- 1012 Rockstrom, J., Steffen, W., Noone, K., Persson, A., Chapin, F.S., Lambin, E.F., Lenton, T.M.,
1013 Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C.A., Hughes, T., van der
1014 Leeuw, S., Rodhe, H., Sorlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M.,
1015 Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K.,
1016 Crutzen, P., Foley, J.A., 2009. A safe operating space for humanity. *Nature* 461, 472–475.
1017 <https://doi.org/10.1038/461472a>
- 1018 Rööös, E., Ekelund, L., Tjärnemo, H., 2014. Communicating the environmental impact of meat
1019 production: Challenges in the development of a Swedish meat guide. *J. Clean. Prod.* 73, 154–
1020 164. <https://doi.org/10.1016/j.jclepro.2013.10.037>
- 1021 Sachs, J.D., 2012. From millennium development goals to sustainable development goals. *Lancet* 379,
1022 2206–2211. [https://doi.org/10.1016/S0140-6736\(12\)60685-0](https://doi.org/10.1016/S0140-6736(12)60685-0)
- 1023 Sadok, W., Angevin, F., Bergez, J.É., Bockstaller, C., Colomb, B., Guichard, L., Reau, R., Doré, T.,
1024 2008. Ex ante assessment of the sustainability of alternative cropping systems: Implications for
1025 using multi-criteria decision-aid methods - A review. *Sustain. Agric.* 753–767.
1026 https://doi.org/10.1007/978-90-481-2666-8_46
- 1027 Sadok, W., Angevin, F., Bergez, J.E., Bockstaller, C., Colomb, B., Guichard, L., Reau, R., Messéan,
1028 A., Doré, T., 2009. MASC, a qualitative multi-attribute decision model for ex ante assessment of
1029 the sustainability of cropping systems. *Agron. Sustain. Dev.* 29, 447–461.
1030 <https://doi.org/10.1051/agro/2009006>
- 1031 Sala, S., Ciuffo, B., Nijkamp, P., 2015. A systemic framework for sustainability assessment. *Ecol.*
1032 *Econ.* 119, 314–325. <https://doi.org/10.1016/j.ecolecon.2015.09.015>
- 1033 Sands, G.R., Podmore, T.H., 2000. A generalized environmental sustainability index for agricultural
1034 systems, *Ecosystems and Environment*.
- 1035 Santiago-Brown, I., Metcalfe, A., Jerram, C., Collins, C., 2015. Sustainability assessment in wine-

- 1036 grape growing in the New World: Economic, environmental, and social indicators for
1037 agricultural businesses. *Sustain.* 7, 8178–8204. <https://doi.org/10.3390/su7078178>
- 1038 Schader, C., Grenz, J., Meier, M.S., Stolze, M., 2014. Scope and precision of sustainability assessment
1039 approaches to food systems. *Ecol. Soc.* 19. <https://doi.org/10.5751/es-06866-190342>
- 1040 Scholes, R.J., Reyers, B., Biggs, R., Spierenburg, M.J., Duriappah, A., 2013. Multi-scale and cross-
1041 scale assessments of social-ecological systems and their ecosystem services. *Curr. Opin.*
1042 *Environ. Sustain.* 5, 16–25. <https://doi.org/10.1016/j.cosust.2013.01.004>
- 1043 Sikdar, S.K., 2003. Sustainable development and sustainability metrics. *AIChE J.*
1044 <https://doi.org/10.1002/aic.690490802>
- 1045 Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, A.K., 2012. An overview of sustainability assessment
1046 methodologies. *Ecol. Indic.* 15, 281–299. <https://doi.org/10.1016/j.ecolind.2011.01.007>
- 1047 Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, A.K., 2009. An overview of sustainability assessment
1048 methodologies. *Ecol. Indic.* 9, 189–212. <https://doi.org/10.1016/j.ecolind.2008.05.011>
- 1049 Smith, C.S., McDonald, G.T., Thwaites, R.N., 2000. TIM: Assessing the sustainability of agricultural
1050 land management. *J. Environ. Manage.* 60, 267–288. <https://doi.org/10.1006/jema.2000.0384>
- 1051 Smyth, A.J., Dumanski, J., 1995. A framework for evaluating sustainable land management. *Can. J.*
1052 *Soil Sci.* 75, 401–406. <https://doi.org/10.4141/cjss95-059>
- 1053 Stanley, E.H., Powers, S.M., Lottig, N.R., Buffam, I., Crawford, J.T., 2012. Contemporary changes in
1054 dissolved organic carbon (DOC) in human-dominated rivers: Is there a role for DOC
1055 management? *Freshw. Biol.* 57, 26–42. <https://doi.org/10.1111/j.1365-2427.2011.02613.x>
- 1056 Steffen, W., Richardson, K., Rockstrom, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R.,
1057 Carpenter, S.R., de Vries, W., de Wit, C.A., Folke, C., Gerten, D., Heinke, J., Mace, G.M.,
1058 Persson, L.M., Ramanathan, V., Reyers, B., Sorlin, S., 2015. Planetary boundaries: Guiding
1059 human development on a changing planet. *Science (80-.)*. 347, 736–+.
1060 <https://doi.org/10.1126/science.1259855>
- 1061 Therond, O., Duru, M., Roger-Estrade, J., Richard, G., 2017. A new analytical framework of farming
1062 system and agriculture model diversities. A review. *Agron. Sustain. Dev.* 37.
1063 <https://doi.org/10.1007/s13593-017-0429-7>
- 1064 Trivino-Tarradas, P., Gomez-Ariza, M.R., Basch, G., Gonzalez-Sanchez, E.J., 2019. Sustainability
1065 assessment of annual and permanent crops: The INSPIA model. *Sustain.* 11, 1–21.
1066 <https://doi.org/10.3390/su11030738>
- 1067 Tsonkova, P., Quinkenstein, A., Böhm, C., Freese, D., Schaller, E., 2014. Ecosystem services
1068 assessment tool for agroforestry (ESAT-A): An approach to assess selected ecosystem services
1069 provided by alley cropping systems. *Ecol. Indic.* 45, 285–299.
1070 <https://doi.org/10.1016/j.ecolind.2014.04.024>
- 1071 Tzilivakis, J., Lewis, K.A., 2004. The development and use of farm-level indicators in England.
1072 *Sustain. Dev.* 12, 107–120. <https://doi.org/10.1002/sd.233>
- 1073 Tzilivakis, J., Warner, D.J., Green, A., Lewis, K.A., Angileri, V., 2016. An indicator framework to
1074 help maximise potential benefits for ecosystem services and biodiversity from ecological focus
1075 areas. *Ecol. Indic.* 69, 859–872. <https://doi.org/10.1016/J.ECOLIND.2016.04.045>
- 1076 Uhlman, B., Laginess, T., Schoeneboom, J., Saling, P., Grosse-Sommer, A., Paczkowski, N., van
1077 Gelder, R., 2016. BASF's AgBalance Methodology.
- 1078 Ullah, A., Perret, S.R., 2014. Technical- and environmental-efficiency analysis of irrigated cotton-
1079 cropping systems in Punjab, Pakistan using data envelopment analysis. *Environ. Manage.* 54,
1080 288–300. <https://doi.org/10.1007/s00267-014-0300-4>

- 1081 Urruty, N., Tailliez-Lefebvre, D., Huyghe, C., 2016. Stability, robustness, vulnerability and resilience
 1082 of agricultural systems. A review. *Agron. Sustain. Dev.* 36, 1–15.
 1083 <https://doi.org/10.1007/s13593-015-0347-5>
- 1084 Van Cauwenbergh, N., Biala, K., Bielders, C., Brouckaert, V., Franchois, L., Garcia Ciudad, V.,
 1085 Hermy, M., Mathijs, E., Muys, B., Reijnders, J., Sauvenier, X., Valckx, J., Vanclooster, M., Van
 1086 der Veken, B., Wauters, E., Peeters, A., 2007. SAFE-A hierarchical framework for assessing the
 1087 sustainability of agricultural systems. *Agric. Ecosyst. Environ.* 120, 229–242.
 1088 <https://doi.org/10.1016/j.agee.2006.09.006>
- 1089 Van Der Werf, H.M.G., Petit, J., 2002. Evaluation of the environmental impact of agriculture at the
 1090 farm level: A comparison and analysis of 12 indicator-based methods. *Agric. Ecosyst. Environ.*
 1091 93, 131–145. [https://doi.org/10.1016/S0167-8809\(01\)00354-1](https://doi.org/10.1016/S0167-8809(01)00354-1)
- 1092 Vanham, D., Leip, A., Galli, A., Kastner, T., Bruckner, M., Uwizeye, A., van Dijk, K., Ercin, E.,
 1093 Dalin, C., Brandão, M., Bastianoni, S., Fang, K., Leach, A., Chapagain, A., Van der Velde, M.,
 1094 Sala, S., Pant, R., Mancini, L., Monforti-Ferrario, F., Carmona-Garcia, G., Marques, A., Weiss,
 1095 F., Hoekstra, A.Y., 2019. Environmental footprint family to address local to planetary
 1096 sustainability and deliver on the SDGs. *Sci. Total Environ.* 693.
 1097 <https://doi.org/10.1016/j.scitotenv.2019.133642>
- 1098 Vihervaara, P., Rönkä, M., Walls, M., 2010. Trends in ecosystem service research: Early steps and
 1099 current drivers. *Ambio* 39, 314–324. <https://doi.org/10.1007/s13280-010-0048-x>
- 1100 Walker, J., 2002. *Environmental Indicators and Sustainable Agriculture*.
- 1101 Wang, L., Li, L., Cheng, K., Pan, G., 2019. Comprehensive evaluation of environmental footprints of
 1102 regional crop production: A case study of Chizhou City, China. *Ecol. Econ.* 164.
 1103 <https://doi.org/10.1016/j.ecolecon.2019.106360>
- 1104 Wang, X., Liu, W., Wu, W., 2009. A holistic approach to the development of sustainable agriculture:
 1105 Application of the ecosystem health model. *Int. J. Sustain. Dev. World Ecol.* 16, 339–345.
 1106 <https://doi.org/10.1080/13504500903106675>
- 1107 Wohlfahrt, J., Colin, F., Assaghir, Z., Bockstaller, C., 2010. Assessing the impact of the spatial
 1108 arrangement of agricultural practices on pesticide runoff in small catchments: Combining
 1109 hydrological modeling and supervised learning. *Ecol. Indic.* 10, 826–839.
 1110 <https://doi.org/10.1016/j.ecolind.2010.01.004>
- 1111 Wohlin, C., 2014. Guidelines for snowballing in systematic literature studies and a replication in
 1112 software engineering. *ACM Int. Conf. Proceeding Ser.* <https://doi.org/10.1145/2601248.2601268>
- 1113 Wustenberghs, H., Coteur, I., Debruyne, L., Marchand, F., 2015. Survey of Sustainability Assessment
 1114 Methods.
- 1115 Yegbemey, R., Yabi, J., Dossa, C., Bauer, S., Nérice Yegbemey, R., Afouda Yabi, J., Sylvestre
 1116 Gerbert Dossa, C., 2014. Novel participatory indicators of sustainability reveal weaknesses of
 1117 maize cropping in Benin 34. <https://doi.org/10.1007/s13593-014-0214-9i>
- 1118 Zahm, F., 2008. IDEA: indicateurs de Durabilité des Exploitations Agricoles. Plate forme d'évaluation
 1119 *Agri-Environnementale* 3, 1–12.
- 1120 Zahm, F., Alonso Ugaglia, A., Barbier, J.-M., Boureau, H., Del'homme, B., Gafsi, M., Gasselin, P.,
 1121 Girard, S., Guichard, L., Loyce, C., Manneville, V., Menet, A., Redlingshöfer, B., 2018.
 1122 Evaluating sustainability of farms: introducing a new conceptual framework based on three
 1123 dimensions and five key properties relating to the sustainability of agriculture. The IDEA method
 1124 version 4. 13th Eur. IFSA Symp. 1–5.
- 1125 Zahm, F., Alonso Ugaglia, A., Boureau, H., Barbier, J., Gasselin, P., Gafsi, M., Guichard, L., Loyce,
 1126 C., Manneville, V., Ugaglia, A.A., 2015. Agriculture et exploitation agricole durables : état de

- 1127 l'art et proposition de définitions revisitées à l'aune des valeurs, des propriétés et des frontières
1128 de la durabilité en agriculture. *Innov. Agron.* 46, 105–125.
- 1129