Available online at www.sciencedirect.com

ScienceDirect

http://www.elsevier.com/locate/biombioe

Indicators of bioenergy-related certification schemes – An analysis of the quality and comprehensiveness for assessing local/regional environmental impacts

*Markus A. Meyer**, *Joerg A. Priess*

Helmholtz Centre for Environmental Research – UFZ, Department of Computational Landscape Ecology, Permoserstr. 15, 04318 Leipzig, Germany

ARTICLE INFO

Article history:
Received 7 July 2013
Received in revised form
17 March 2014
Accepted 18 March 2014
Available online 8 April 2014

Keywords:
Certification schemes
Environmental impact
Bioenergy
Ecosystem services
Criteria and indicators
EU Renewable Energy Directive

ABSTRACT

Bioenergy is receiving increasing attention because it may reduce greenhouse gas emissions, secure and diversify energy supplies and stimulate rural development. The environmental sustainability of bioenergy production systems is often determined through life-cycle assessments that focus on global environmental effects, such as the emission of greenhouse gases or air pollutants. Local/regional environmental impacts, e.g., the impacts on soil or on biodiversity, require site-specific and flexible options for the assessment of environmental sustainability, such as the criteria and indicators used in bioenergy certification schemes.

In this study, we compared certification schemes and assessed the indicator quality through the environmental impact categories, using a standardized rating scale to evaluate the indicators. Current certification schemes have limitations in their representation of the environmental systems affected by feedstock production. For example, these schemes predominantly use feasible causal indicators, instead of more reliable but less feasible effect indicators. Furthermore, the comprehensiveness of the depicted environmental systems and the causal links between human land use activities and biophysical processes in these systems have been assessed. Bioenergy certification schemes seem to demonstrate compliance with underlying legislation, such as the EU Renewable Energy Directive, rather than ensure environmental sustainability. Beyond, certification schemes often lack a methodology or thresholds for sustainable biomass use. Lacking thresholds, imprecise causal links and incomplete indicator sets may hamper comparisons of the environmental performances of different feedstocks. To enhance existing certification schemes, we propose combining the strengths of several certification schemes with research-based indicators, to increase the reliability of environmental assessments.

* *Corresponding author*. Tel.: +49 3412351625; fax: +49 3412351939.
E-mail address: markus.meyer@ufz.de (M.A. Meyer).
http://dx.doi.org/10.1016/j.biombioe.2014.03.041

Abbreviations

CSBP	Council on Sustainable Biomass Production
C&Is	criteria and indicators
DPSIR	driving forces – pressures – states – impacts – responses
ESS	ecosystem services
EU RED	EU Renewable Energy Directive
FSC	Forest Stewardship Council
GBEP	Global Bioenergy Partnership
GlobalGAP	Global Good Agricultural Practice
GGL	Green Gold Label
IWPB	Initiative Wood Pellet Buyers
ISCC	International Sustainability and Carbon Certification
LU/LUC	land use and land-use change
NTA	Netherlands Technical Agreement
PEFC	Programme for the Endorsement of Forest Certification
RSB	Roundtable on Sustainable Biomaterials
SAN	Sustainable Agriculture Network
SEM	standard error of the mean
SOC	soil organic carbon
SFI	Sustainable Forestry Initiative

1. Introduction

Bioenergy is receiving increasing attention because it is assumed to be associated with the following major advantages over fossil fuels [1-4]:

- Reduction of greenhouse gas (GHG) emissions and strengthening of the environmental sustainability of energy provision
- Securing and diversifying the energy supply
- Positive socioeconomic impacts such as increased energy access in developing and jobs in developed countries

The arguments in favor of bioenergy can be summarized under the concept of sustainability as defined by the Brundtland Commission [5]. The aspects listed above show that several dimensions of sustainability are of importance, namely the economic, environmental and social dimensions [6]. According to neoclassical theory, economic sustainability is ensured through market mechanisms [7]. Environmental and social sustainability are often not ensured through these mechanisms and require government interventions, for example, quotas for bioenergy or subsidies to overcome market failures [8]. Even if environmental and social sustainability are considered for bioenergy, Robbins [9] stated that it is currently unclear how to assess the sustainability of bioenergy from both environmental and socioeconomic perspectives.

The major environmental impact categories of bioenergy feedstock production have been summarized to GHG emissions, air pollutants, soil quality, water quality, water availability or quantity, biodiversity and land-use and land-use change (LU/LUC) based on scientific literature [10–13] and broader stakeholder panels [14]. To a great extent, the environmental sustainability of bioenergy production systems is evaluated with well established life-cycle assessments (LCAs), assessing large-scale or globally occurring environmental effects, such as GHG emissions or air pollutants, along the major steps of the supply chain [10,15]. The highly site-specific and locally/regionally occurring environmental impacts of feedstock production in the first step of most of the bioenergy supply chains are difficult to assess in LCAs. Impacts on soil quality, biodiversity and land use change, water availability and water quality [16,17] are often insufficiently covered. These limitations comprise necessary but missing regional thresholds to ensure the stability of the ecological system. Such thresholds are not easily integrated into highly standardized LCAs. Existing LCAs assessing environmental impacts often disregard the interaction for example between different regulating ecosystem services (ESS) and biodiversity, such as the buffering capacity of environmental impacts of agriculture or forestry [18,19]. In the context of bioenergy feedstocks and sustainability, this type of assessment of interactions is supposed to extend the EU RED, i.e., the provision of "basic ecosystem services" such as erosion control should be accounted for if biomass is produced for bioenergy [20]. Dale et al. [21] recommend to determine water quality and soil quality impacts of bioenergy feedstock production in addition to LCAs, e.g., nutrient export to water bodies or soil loss. A regional water quality assessment will more likely allow to determine, whether regional thresholds of nutrient exports that ensure good ecological status of water bodies are met.

Site-specific and flexible options for the assessment of local/regional environmental impacts and other aspects of sustainability could be sets of criteria and indicators (C&Is) as used in certification schemes. Such a site-dependent audit approach allows assessing the environmental impacts and their interactions mentioned above. C&Is are currently under development or are at an early stage of implementation for bioenergy but have been extensively applied for a longer period to other products from forestry or agriculture. Examples of C&Is are the Forest Stewardship Council (FSC) for timber or the Sustainable Agriculture Network (SAN) as a label for Good Agricultural Practices [2]. Especially FSC provides nationally or regionally adapted indicator sets [22]. Several bioenergy certification schemes are used to demonstrate compliance with the EU Renewable Energy Directive 2009/28/EC (EU RED) [23].

Despite the common aim of EU RED compliance for most of the bioenergy schemes, an increasing number of alternative schemes may contribute to confuse stakeholders and decrease the acceptance of certification schemes in general [24,12]. On the one hand, comprehensive and clearly defined requirements may exclude producer groups [2], e.g., in developing countries, and augment certification costs due to increasing effort, such as audits. On the other hand, vaguely defined and less comprehensive schemes may allow for a higher market penetration, but more likely disregard major environmental or social impacts and are not acknowledged by NGOs [25,26]. An increase in EU imports of biomass for bioenergy might induce or enhance deforestation in countries with prevailing primary forests [27] and the need to export goods. Thus, overexploitation is more likely to occur in developing countries than in developed countries. To avoid or abate e.g., deforestation, a set of C&Is must be agreed upon internationally to cover international biomass trade [28]. International criteria might exceed the local requirements for bioenergy sustainability or might set foci other than the locally intended ones [29]; e.g., criteria might focus on environmental aspects in developed countries, such as sequestering carbon or halting biodiversity loss instead of ensuring food security in developing countries [13]. Such potential discrepancies may provide additional obstacles for implementation.

Beyond existing reviews [2,12,13,26,29], this paper, assesses the comprehensiveness and quality of indicators used by bioenergy, forestry and agricultural certification schemes. Against the background of conflicting goals for bioenergy certification discussed above, we develop and apply standardized rating scales for indicators grouped into six environmental impact categories to identify their reliability and feasibility. We focus on local/regional environmental impacts, which require site-specific information, affect predominately the local/regional environment and are usually not covered by LCAs. Beyond rating the individual indicators, certification schemes are evaluated at the scheme level based on the ESS cascade [30] to analyze their comprehensiveness and the quality of the representation of the potentially affected environmental system. The aim is to test whether certification schemes are able to show trade-offs between biomass use and other ecosystem services.

2. Material and methods

2.1. *Selection of certification schemes and indicator sets*

In this paper, indicator sets for certification have been selected for evaluation. We used sets from bioenergy, agriculture and forestry. The latter two have the advantage of a much longer lasting application of C&Is. Concentrating on the currently rather limited number of specific schemes for bioenergy would have led to a very small set of C&Is, ignoring relevant and important C&Is applied in related sectors.

First, the EU might consider the extension of bioenergy specific with forestry schemes as a relevant policy option for solid biomass for bioenergy in the EU, e.g., by using additional forestry indicators for sustainability certification [31]. Therefore, an evaluation of studies is conducted, assessing the environmental impacts of forest management with a focus on bioenergy production. To identify major characteristics of forestry certification schemes, we selected the FSC and the Sustainable Forestry Initiative (SFI), a major scheme of the meta-standard “Programme for the Endorsement of Forest Certification” (PEFC), which are globally dominating and largely applied certification schemes in forestry [2,32]. We avoided meta-standards since they typically do not have indicators sets for the actual environmental assessment.

Secondly, new technologies to enhance the transport, storage and co-firing characteristics, such as torrefaction, are under development. These technologies might create additional feedstock options, for instance agricultural residues, such as straw, shells and others, which currently may be used to a limited extent [33]. Therefore, overarching and globally applied agricultural certification schemes, i.e., SAN and Global Good Agricultural Practice (GlobalGAP), are needed to cover feedstocks not targeted by bioenergy certification schemes, predominately aiming at selected bioenergy crops. The relevance of agricultural certification schemes shows NTA 8080 and other bioenergy certification schemes as they use agricultural certification schemes, which we also selected in this paper, to ensure compliance with environmental sustainability requirements [13]. Despite the fact that GBEP is no operational certification scheme, we included it in our assessment since its indicator set reflects the consensus of numerous governments and international institutions and because it is a framework to assess bioenergy sustainability [12].

2.2. *Requirements and rating scales for indicator evaluation*

The major requirements for indicators are reliability and conceptual soundness, feasibility, i.e., measurability and practicality, and relevance for the end user [2,34–36]. The requirements for an indicator discussed in this section are rated on a five step scale. Bockstaller et al. [34] have demonstrated the methodological suitability of such an approach at the indicator level by evaluating sets of agri-environmental indicators for crop production and farming systems, which are methodologically comparable to the certification scheme indicators evaluated in this paper.

We rate the individual indicators for feasibility in three requirement subcategories and for reliability in four requirement subcategories, two exemplary requirement subcategories each are listed in Table 1 and the remaining ones in Appendix A.

The first rated subcategory for reliability is the *Indicator type* [34,37]. For practical implementation, we followed the logic of the **D**riving forces – **P**ressures – **S**tates – **I**mpacts – **R**esponses

Table 1 – *(upper part)* Rating scale for the reliability of indicators, subcategory *Indicator type* adapted from Bockstaller et al. [34]; *(lower part)* rating scale for the feasibility of indicators, subcategory *Required resources (assessment interval)*

Indicator type (cause vs. effect-related)	
1 Driver	Management practice
2 Driver	Management practices related to state or impact
3 Pressure	Release of pollutants or sediment
4 State	Concentration of pollutant in environmental compartment
5 Impact	Environmental changes attributable to pollutants or sediments
Required resources (assessment interval)	
1 Daily assessment/measurements required	
2 Seasonal assessment/measurements required	
3 Annual assessment/measurements required	
4 Less than annual measurements	
5 No measurement, only completing a survey	

(DPSIR) framework of the European Environment Agency [36], extending preceding frameworks, such as the **P**ressure–-**S**tate–**R**esponse framework, applied by the OECD and the UN [38]. We present an application example for the DPSIR framework for rising wood pellet demand, conceptually based on Bockstaller et al. [39] and Svarstad et al. [38]. A rising demand of wood pellets may require to apply more fertilizer for shorter rotation cycles of forest plantations, e.g., *Pinus* spp., (Driving force). Consequently, increased fertilizer application may increase the nutrient runoff to surface water bodies (Pressure), which may lead to higher nutrient concentrations (State), i.e., possibly eutrophication, which may change e.g., the species composition (Response). Thus, an indicator of an environmental pressure such as the nutrient load from pine plantations on a water body would be rated as "three" on the five step scale, and a state indicator such as the nutrient concentration in a river would be rated as "four" or the nutrient application rate in the driver category as "one". The closer the assessment is to the environmental impact, the more information on the environmental impact is expected to be considered. The second subcategory for reliability is the *Validity of indicators*. We rate the validity, according to a rating scale, see Table A.1 in Appendix A, modified from Bockstaller et al. [34], which has been developed by Bockstaller and Girardin [40]. We rate the indicators (i) based on scientific literature, i.e., whether peer-reviewed articles use and confirm the exact indicator (value 4), whether the indicator is under debate in the scientific literature (value 3), only confirm the calculation method of the indicator or even reject the indicator (value 2). (ii) Other options are that the indicator needs to agree with locally collected data (value 5) or is typically gained from a validated model (value 4), a partly or only regionally validated model (value 2). If no validation is possible due to the rating in the subcategory *Indicator type* rated as given for indicators on management practices (value 1 or 2), we rate the indicator with a value of "three". The third subcategory for reliability is the *Response time* since an immediate response or a reponse at least in the time frame of political decision making [10,36] enable timely detection and counteraction to the expected or observed environmental problems. We rate the response time of indicators based on peer-reviewed publications.

The first subcategory for feasibility is the *Data requirement*, assessing the ease of data access [2,34,36,39]. We rate indicators based on (i) the nature of the data, i.e., whether it can be obtained from authorities or other data sources (value 5), requires questioning the feedstock producer (value 4) or measurements are required (value 1–3). (ii) The measurement scale is additionally used for the rating [41], i.e., whether indicator data has to be measured at each field or farm individually (value 1) or whether one regional assessment is sufficient for the indicator (value 3). In addition, indicators may be attributed to the field/farm or the regional scale depending on the individual case (value 2), e.g., influenced by farm size (group certification) or an imprecise definition of the indicator in the certification scheme. The second subcategory for feasibility is the *Qualification requirement* [39,34,2] covering the ease or difficulty to assess an indicator due to its specificity or the required expert knowledge (requirements defined in Appendix A). High qualification requirements may be an obstacle for small scale producers, especially in developing countries [24]. The third subcategory for feasibility is the *Required resources (assessment interval)*, i.e., the frequency of possible measurements influences the effort and costs for certification. The fourth subcategory for feasibility is *Clearly defined thresholds*. We rate the existence of target values, reference conditions or thresholds because their availability influences the measurability [11]. A threshold or a possible source to derive it provided by the scheme facilitates the interpretation of feedstock impacts regarding sustainability during the auditing process [41].

The relevance of an indicator first depends on its acceptance by stakeholders, i.e., whether the indicator is suitable to address a certain environmental impact category [36], and secondly on the degree to which stakeholders are involved in the selection process [26]. Data on the preferences of stakeholders is only available for criteria or for the even higher aggregation level of environmental impact categories, but is not available for the corresponding indicators (c.f. Buchholz et al. [35]). The lack of data might also be due to the fact that the development and choice of the rather technical indicators are related to the expertise of the practitioners or scientists. Therefore, the relevance of the indicators cannot be rated but will be checked indirectly by its fit to the relevant environmental impact categories.

We rate indicators that provide direct information about the occurrence or avoidance of environmental impacts. The indicators are aggregated by local/regional environmental impact category on a composite scale. In this context, a composite scale is the combination of several indicators into a thematic category, i.e., we compute the arithmetic mean of all indicators per certification scheme per environmental impact category and the indicator subcategories respectively. Similarly, the standard error of the mean (SEM) is calculated to assess the uncertainty of the arithmetic mean. We assess the indicator sets for the environmental impact categories soil quality, water quality, water availability or quantity, biodiversity and LU/LUC. Soil quality indicators cover indicators on both the management of soils and soil properties. Water quality and availability indicators assess both management activities with an impact on water bodies as well as state indicators of water bodies. Biodiversity indicators may assess the state of conservation areas, species composition or management activities for biodiversity. LU/LUC indicators give information on characteristics of a land use, e.g., carbon payback time, or assess whether no-go areas according to the EU RED definition have been converted for bioenergy feedstocks. The composite scale *Other* comprises indicators without a link to the listed environmental impact categories, which are related to the environmental stability of a system such as indicators on sustainable harvest levels. If applicable, indicators are attributed to two composite scales if a clear link to both is given, e.g., "no conversion of areas of high conservation value" to biodiversity and LU/LUC or "no removal of coarse woody debris" to soil quality and biodiversity.

Internal consistency is ensured by excluding indicators that do not directly measure environmental impacts, i.e., contextual knowledge is used according to Coste et al. [42]. Background knowledge on the environmental indicators, e.g., given by the certification scheme, allows to categorize the

Table 2 – Number of indicators analyzed for each scheme and each environmental impact category (=composite scale) and abundance of aspects in certification schemes excluded from evaluation to ensure internal consistency of composite scales; these results are based on CSBP [43], GBEP Task Force [14], GGL [44], GlobalGAP [45], ISCC [46], IWPB [47], Netherlands Standardization Institute [48,49], REDcert [50], RSB [51], SAN [52] and forestry [29,32,53–56]. For GGL, the agricultural source criteria (GGL2) are assessed.

		GBEP	NTA8080	ISCC	RED_{cert}	GGL	RSB	CSBP	IWPB	SAN	GlobalGAP	Forestry
Composite scales												
Total	*87*											
Soil quality	31	1	9	11	2	0	5	5	8	5	2	30
Water quality	17	2	4	6	7	4	3	6	3	4	7	6
Water availability	9	3	2	2	0	1	1	2	3	1	2	6
Biodiversity	18	3	5	1	1	1	7	3	1	10	2	13
LU/LUC	9	5	3	2	1	0	1	0	3	1	1	1
Others	3	2	0	0	0	0	0	0	1	0	0	3
Abundance of excluded aspects												
Off-site handling rules and machinery maintenance (e.g., disposal of plant protection product containers)	46	1	1	21	7	1	0	0	0	3	12	0
Demonstration of compliance with existing legislation or other rules such as certification schemes, manuals or rules (e.g., registration of product use)	32	0	3	5	3	0	4	3	6	5	3	0
Management plan or other unspecified action or goal required	33	0	0	2	1	2	19	4	0	2	3	0
Qualification and training of staff	10	0	0	2	1	0	0	1	3	1	2	0
Generic monitoring (e.g., soil quality has to be assessed)	6	0	0	0	0	2	2	0	0	2	0	0

indicators. Internal consistency is required since the arithmetic mean should only be calculated for indicators that measure the same latent variable, i.e., environmental impact category. We exclude indicators, for example, if they assess whether legislation is covering environmental impacts, e.g., on water quality. In this case, certification schemes assume that environmental impacts are avoided (complying with existing regulations).

We list the indicators we included and excluded for each scheme in Table 2.

2.3. *The ecosystem service cascade for evaluation of certification schemes*

Assessing certification schemes by only looking at indicators individually would disregard the schemes' quality and comprehensiveness concerning the use of environmental systems and the services/disservices derived thereof. A widely accepted concept to determine and quantify the human use of the environment is ESS [57,58].

The ESS cascade [30] is a conceptual framework used to connect ESS to the underlying ecosystem structures and processes and to the human benefits derived from the use of the ecosystem. Ecosystem structures and processes are the basis to derive thresholds for the sustainable provision of an ESS [30,57], i.e., the ecosystem capacity. For example, the ecosystem capacity can be used to answer questions about the critical limits or thresholds [59] for e.g., the extraction of tree biomass to sustain forest stocks . Because this evaluation focuses on local/regional environmental impacts, it is beyond our scope to depict the socioeconomic components of the ESS cascade, i.e., the human benefits and (monetary) values. We focus on biophysical and ecological structures and functions and their alteration due to the use of ESS. The ecological and the socioeconomic systems are linked by the use of ESS [60], e.g., biomass use. In practice, the ESS cascade has been used as a conceptual framework to embed indicators of different provisioning services, e.g., biomass production [61,62], and regulating services, e.g., water purification [63], of the underlying environmental systems. In addition, the ESS cascade has also been used to visualize the interaction of indicators within and between the different components of the ESS cascade [62,64]. Maes et al. [63] and Van Oudenhoven et al. [62] add land management to the beforehand mentioned components of the ESS cascade. The necessity of including land management was previously stated by Haines-Young and Potschin [30] but was not implemented. Like Ojima et al. [65], we included land management aspects because indicators of ESS describe the use of natural capital but do not provide insight into the extent that the use of ESS is altered by human land use activities, i.e., agricultural practices such as irrigation or fertilization or conservation measures such as field margins for biodiversity.

In this study, we use the term "human land use activity" because this term includes land management, land conversion and changes in the structure of the landscape [66]. Therefore, indicators of human land use activities enable the assessment of the intensity of land use associated with different types of and options for biomass provision. For example, changes in production practices or landscape planning are likely to affect ecosystems, i.e., the structures, processes and capacity. A better representation of the interaction

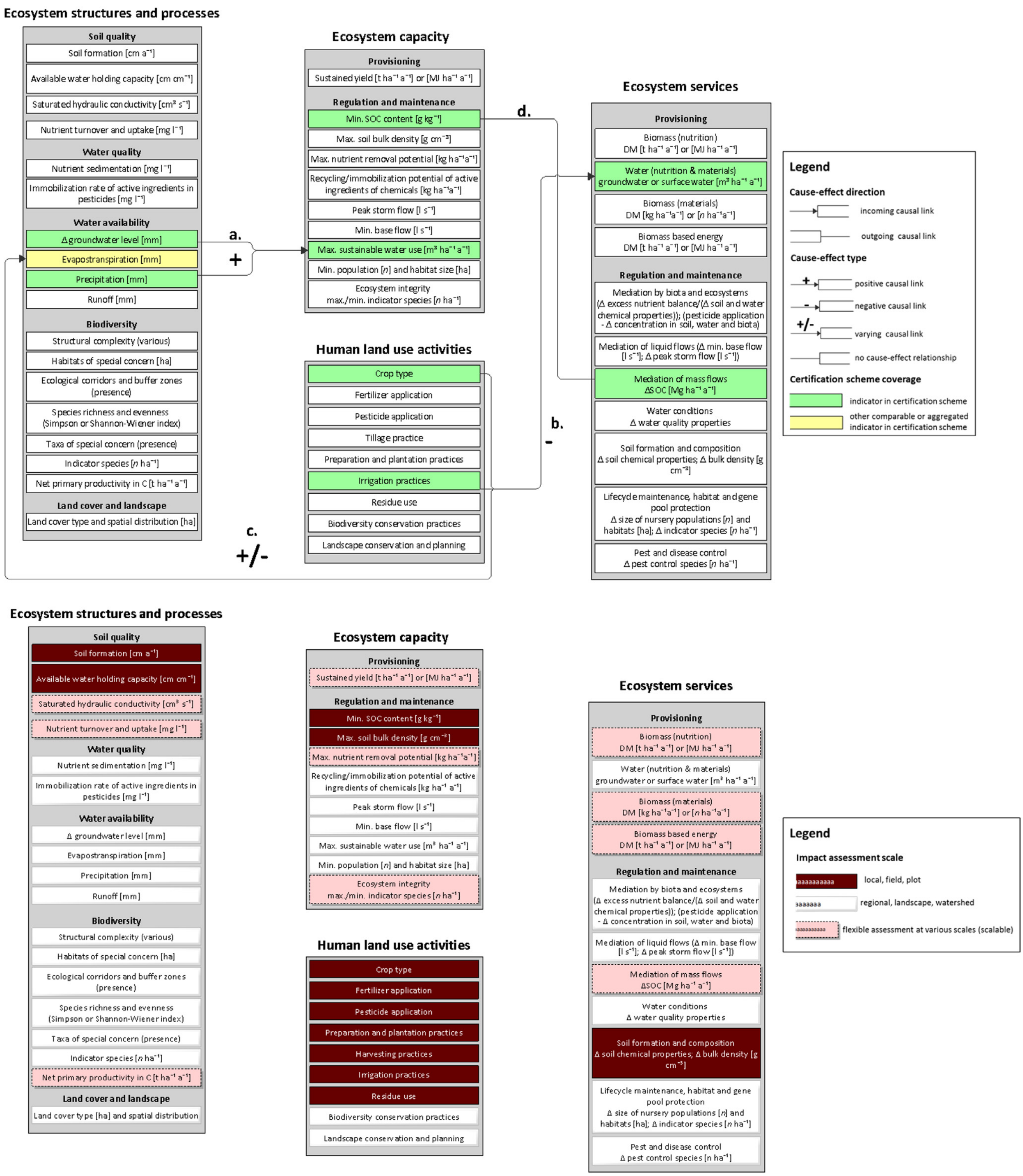

Fig. 1 – *(upper part)* ESS cascade (modified from CICES [67], Maes et al. [63], Potschin and Haines-Young [60], Van Oudenhoven et al. [62]) as an analytical framework to evaluate certification schemes for bioenergy feedstock production; the components shown are ecosystem structures and processes (underlying biophysical mechanisms), ecosystem capacity (sustainability thresholds for ESS use) and ESS (actual use of ESS or creation of disservices). The arrows indicate a. positive, b. negative, c. varying and d. no causal link. The selected indicators are adapted to the major impacts of bioenergy production identified from Dale and Beyeler [68], De Groot et al. [57], Haines-Young and Potschin [30], Kandziora et al. [64], Kienast et al. [69], Lattimore et al. [53], McBride et al. [11], McElhinny et al. [70], Schoenholtz et al. [55], Wascher [71]. *(lower part)* Spatial impact assessment scales of the ESS cascade adapted for bioenergy feedstock production. The impact assessment scales are generally based on De Groot et al. [57] and Efroymson et al. [10] and are specifically based on Sposito [72] for hydrology and Turner et al. [73] for landscape patterns.

of human land use activities, ecosystems and ESS use might help to identify environmentally especially harmful biomass use and land management practices. More reliable results could allow decision makers to better target, e.g., mitigation activities.

In this study, the ESS cascade is extended from a conceptual to an analytical framework for bioenergy feedstock production (Fig. 1). The ESS cascade is converted and expanded into an analytical tool to assess the quality of certification schemes. The latter are implemented within the framework to assess the sustainability of feedstock provision with environmental C&Is; i.e., the adverse environmental impacts should be revealed to facilitate mitigation or avoidance as requested by Van Dam et al. [13]. Thus, the extended ESS cascade is applied to investigate whether certification schemes represent biophysical processes for feedstock production in a qualitatively and quantitatively useful manner. We apply the widely used "Common International Classification of ESS – CICES" v4.3 [67], which has undergone several rounds of international review and consultation, to ensure assessing all major ESS, which may be affected by bioenergy feedstock production.

The mapping used for the certification scheme indicators is presented in Fig. 1. For the different certification schemes we analyzed, we focused especially on the representation of causal links and the coverage of ecosystem structures and functions represented in the extended ESS cascade, i.e., the quality of the representation of the environmental system. For example, does a certification scheme include indicators that would reveal if biomass use affected other ecosystem services such as surface or groundwater provision? Does a certification scheme include the link from fertilized pine plantations to a possible ground- or surface-water pollution and does it provide the relevant indicators on, e.g., water quality and fertilization practices? We took the individual indicators per certification scheme, related them to the environmental system and indicated the causal links and components covered.

For an overview, we counted the actual number of indicators for each of the four components of the ESS cascade displayed in Fig. 1 and rated them on a three step scale based on thirds. For causal links, the certification schemes are compared with their peers. The certification scheme with the highest number of causal links has the best rating, i.e., 100%, and is used as a benchmark and rated as done for the indicators. The indicators and causal links for each scheme are displayed in Appendix A.

The following three types of common causal links and links without cause–effect relationships are found in the evaluated certification schemes and indicator sets:

a. **Positive causal link (<u>In</u>crease in X causes an <u>in</u>crease in Y):**

Example. "The participating operator provides objective evidence demonstrating that her/his/its biomass/biofuels operation(s) does/do not contribute to exceeding the replenishment capacity of the water table(s) [...]," RSB [51]. This statement implies that the maximal sustainable water use does not negatively affect the groundwater table and is adapted to the local level of precipitation. Therefore, both a higher precipitation and a higher change of the groundwater table, i.e., a lower decline, may result in a higher maximal sustainable water use.

b. **Negative causal link (<u>In</u>crease in X causes a <u>de</u>crease in Y):**

Example. The feedstock provider measures the water use per area and uses irrigation techniques that conserve water most, e.g., CSBP [43]. In other words, if more irrigation techniques with low water use are applied (replacing inefficient technologies), the use of water units per unit bioenergy feedstock will decrease per ha.

c. **Varying causal link (<u>In</u>crease in X causes an <u>in</u>crease or <u>de</u>crease in Y):**

Example. "Have systematic methods of prediction been used to calculate the water requirement of the crop?" GlobalGAP [45]. Options for actions are suggested in the explanation of the indicator. The actions may be operationalized as follows: The amount of water used varies with the crop type. Hydrologically, the upward flux of water via plants and soil is termed evapotranspiration. The choice of a crop may increase or decrease evapotranspiration. Because this biophysical flux is not named in the indicator, but is only implicitly considered, it is highlighted in yellow.

d. **No cause–effect relationship**: The soil organic carbon content is maintained or improved, e.g., GBEP Task Force [14]. The definition of the indicator specifies both the ecosystem capacity and the parameter to be measured to determine the ESS use, i.e., mediation of mass flows. Here, a thematic link between ecosystem capacity and ESS is given instead of a cause–effect relationship.

Additionally, we need to assess how certification schemes are able to overcome the challenge of the necessity of assessing (i) environmental impacts at scales beyond the field/farm level [12] and (ii) the interaction and accumulation of environmental impacts beyond different spatial scales [10,37] and how to distribute target values or thresholds [74,75]. Within this study, the relevant spatial scales from both the literature on actual indicators and from specific studies on scales to determine specific environmental parameters are shown in Fig. 1. Because this study focusses on local/regional scale environmental impacts, there are no indicators included beyond those scales. Local scale, also plot or field scale, is typically areas less than 1 km^2 and regional, also landscape or watershed scale ranges from 1 to 10,000 km^2 [37,57]. There are some indicators that are more flexible and provide reasonable results at both of the considered scales. For example, the sustained yield and the underlying primary productivity can be scaled up or down for largely homogenous ecosystems, such as those in forestry, where sustainable harvest levels or wood resources and residues are common indicators [32].

3. Results and discussion

3.1. *Major characteristics of certification schemes*

The major characteristics evaluated in this study are those identified as relevant by existing reviews [10,13,76,77], and the evaluated certification schemes and their indicators are introduced in the following sections.

Table 3 shows that only GBEP, NTA 8080, GGL and CSBP target all types of bioenergy. CSBP intends to certify any type of bioenergy from ligno-cellulosic biomass. ISCC, REDcert and RSB originally were developed to demonstrate compliance with national or supra-national legislation, i.e., the EU RED, which primarily cover biofuels and bioliquids [13]. Currently, these schemes are being partially extended and revised to certify solid and gaseous bioenergy to ensure compliance with regulations in potential new versions of the EU RED. NTA 8080 is also used to demonstrate EU RED compliance for biofuels and bioliquids but is the implementation of the "Testing framework for sustainable biomass," the so-called Cramer Criteria, which originally focused on any type and use of sustainable biofuels and other products from biomass [12]. The remaining certification schemes have been developed to ensure sustainable production of agricultural or timber products. To ensure cost-effectiveness, the EU might consider forest certification schemes to be a proof of sustainable production of solid biomass [31]. Table 3 shows that certification schemes for bioenergy attempt to assess the entire supply chain of a product to demonstrate, for example, the higher environmental sustainability than that of fossil energy carriers. The agricultural or forestry certification schemes are rather purpose specific; for example, the schemes demonstrate low-impact cultivation techniques or sustainable forest management [12] and thus focus on feedstock production rather than on the final product. In the latter aspect they differ from bioenergy certification schemes.

3.2. *Indicator evaluation*

3.2.1. *Overview*

For the requirements for indicators, the mean of the indicators for certification schemes in Fig. 2 shows that most of the certification schemes are rated at the center of the scale at this aggregation level. The mean for the *Required resources (assessment interval)* with an above-average rating and the mean for the *Indicator type* with a below-average rating for most of the schemes deviate from the general tendency toward a centered rating.

The pattern of the *Required resources (assessment interval)* and *Indicator type* may be interpreted as the common trade-off between the feasibility and the reliability of indicators (c.f. Payraudeau and van der Werf [37]).

The thematic abundance of indicators not suitable for a direct environmental assessment and therefore excluded for internal consistency of the composite scales has been shown in Section 2.2 in Table 2. Analyzing such excluded indicators gives insight into how certification schemes aim to demonstrate environmental sustainability without an environmental assessment. The majority of the aspects excluded are those not directly related to biomass cultivation or harvesting but are instead related to the handling of equipment and post-production waste or to the documentation of farming activities. The evaluated certification schemes build on cross-compliance or are at least partly set up as a meta-standard. Indicators assess whether legislation or other certification schemes are fulfilled but do not assess whether the environmental impacts of bioenergy production are addressed. Indicators that require the establishment of management plans or actions to achieve a target, such as maintaining water quality, are equally abundant. In minor abundance is the qualification of staff members conducting different tasks in biomass cultivation and processing and generic monitoring activities, such as those related to soil quality.

This overview may provide the impression that the selection of most of the indicators is predominately driven by the aim to allow for highly feasible or practical and probably cost-effective assessment, e.g., leading to assessments that do not require (on-site) measurements, such as demonstrated compliance with local legislation or the review of existing documentation. The named indirect assessment approaches not only consume less time and fewer resources but also do not require an understanding of environmental processes or measurement techniques for an on-site assessment for either the certified party or for the auditor. Certification schemes that require the establishment of generic management plans or monitoring without any consideration of local environmental conditions and processes may facilitate a worldwide sustainability assessment.

3.2.2. *Evaluation of indicators by requirements and by composite scales*

The overview in Section 3.2.1 revealed that a high aggregation level does not reveal significant differences between certification schemes. Therefore, the results for the ratings of certification scheme indicators are analyzed at the less aggregated level of composite scales and are grouped by the indicator requirements and their subcategories, see Fig. 3.

Based on reliability and conceptual soundness, the *Indicator type* has a nearly universal low rating (value 1–2); i.e., driver indicators on management practices are used, especially for water quality and water availability. Biodiversity and LU/LUC indicators are partially state or impact indicators (value 4–5). These indicators determine whether land use types are converted for biomass production for bioenergy. An example of such state indicators are spatial biodiversity indicators; e.g., there is no bioenergy feedstock production in areas of high conservation value (ecosystems, species). Such indicator demonstrates or intends to demonstrate compliance with EU RED (ISCC, REDcert, IWPB, GGL). For example, the certification schemes named above assess whether areas of high conservation value or of specific land use types with high carbon stocks, such as peatland, are converted for bioenergy feedstock production. Other EU RED compliance demonstrating schemes (NTA 8080, RSB) without such a pattern have

Table 3 – Major characteristics of certification schemes based on BEFSCI [76], CSBP [43], EC [78], FSC [79], GBEP Task Force [14], GGL [44], GlobalGAP [45], ISCC [46], IWPB [47], Netherlands Standardization Institute [48,49], REDcert [50], RSB [51], SAN [52], SFI [80].

	Bioenergy								Agriculture		Forestry	
	GBEP	NTA8080	ISCC	REDcert	GGL[b]	RSB	CSBP	IWPB	SAN	GlobalGAP	FSC	SFI
Major characteristics:												
Applicable biofuel type												
solid	x	x	(x)	(x)	x	(x)	(x)	x			(x)	(x)
liquid	x	x	x	x	x	x	x					
gaseous	x	x	(x)	(x)	x	(x)	(x)					
Spatial scope for application	Global	Global	Global	EU[a]	Global	Global	US	Global	Global	Global	Regional	US/ Canada
EU RED recognition		x	x	x		x						
Degree supply chain coverage[c]	FTPD	FTPD	FTPD	FTPD	FTPD	FTPD	FT	FTPD	F	F	F	F

[a] Few third countries (e.g., Belarus, Ukraine).
[b] Agricultural and forestry source criteria.
[c] Supply chain coverage: feedstock (F), transport (T), processing (P), distribution (D).

indicators other than spatial indicators that address the protection or restoration of ecological corridors or buffer zones. The *Validity of indicators*, with the exceptions of the composite scale for water availability and, more significantly, the composite scale *other* for the non-attributable indicators, could be largely characterized as being validated by models or by agreement in the scientific literature (value 4). The *Response time*, see Fig. 3, of the chosen indicators is typically one to five years or is not measured, as for causal indicators (value 3), i.e., *Indicator type* (value 1 or 2). The latter option is more likely because Fig. 3 shows that most of the indicators are causal. Biodiversity and LU/LUC indicators partially show immediate responses (value 5). The rating pattern for Biodiversity and LU/LUC is comparable to the requirements for *Indicator type* and for the described indicators; see Fig. 3; i.e., the chosen impact indicators are associated with short response times.

Based on the results for feasibility, the *Data requirement* for the evaluated certification schemes shows that indicators for which data is available at other scales (value 3) or which require data from field observations and questionnaires but measurements (value 4) are not predominately used. The *Qualification requirement* greatly varies for the different composite scales. The biodiversity indicators are difficult to assess or require prior knowledge. At the least, general higher education, a university degree in agricultural science, or vocational training is required for the assessment (value 2–3). In contrast, the indicators chosen for water availability, e.g., water use per area, require no education or at least no more than a short introduction (value 4–5). The *Required resources (assessment interval)*, soil quality, water quality and availability and other indicators are assessed predominately at intervals longer than one year (value 4) or do not even require field assessment (value 5). Biodiversity and LU/LUC impacts need to be assessed with a higher frequency; some must be assessed annually (value 3). The comparable patterns for *Data requirement* and *Required resources (assessment interval)* show that the data type and collection mode and the required resources seem to be correlated, i.e., the more effort that data collection for an indicator requires, the higher the frequency of assessment and vice versa. With respect to the requirement *Clearly defined thresholds*, certification schemes mostly only indicate (value 3) how to derive target values/thresholds or use causal indicators. Causal indicators do not require an actual threshold. Instead, the question is whether a (sustainable) management practices is applied or not, i.e., an assessment of compliance or non-compliance. LU/LUC indicators are an exception; for these indicators a threshold is typically given because their formulation implies that there must not be any land conversion for bioenergy feedstock production.

Trade-offs between feasibility (*Data requirement, Required resources (assessment interval)*) and reliability (*Indicator type, Response time*), mentioned in Section 3.2.1, are especially pronounced for the composite scale for water availability but are also pronounced for soil and water quality. For water availability, the requirements characterizing feasibility, *Data requirement* and *Required resources (assessment interval)*, are highly rated (value 4 or 5). The *Data requirement* can be met with field observations or questionnaires (value 4). The *Required resources (assessment interval)* are minimal because only surveys and no measurements need to be conducted (value 5). Because it is only necessary to complete a survey without measurements and this process requires even less assessment effort than the least frequent measurement, personnel resources and equipment can be saved relative to indicators that are regularly measured.

The indicator requirements for reliability are rated low. Driver indicators (management practices) that measure no response for the *Indicator type* (value 1–2) and *Response time* (value 3) are used. Such a trade-off is not pronounced for the *Validity of indicators* and their feasibility (*Data requirement, Required resources (assessment interval)*) because both are often highly rated (value 4). I.e., many driver indicators are either validated by models or are widely accepted in the scientific literature. The latter explanation applies to many of the indicators in this study. The comparable high ratings for the *Data requirement* and *Required resources (assessment interval)* reveal that certification schemes preferably use feasible indicators.

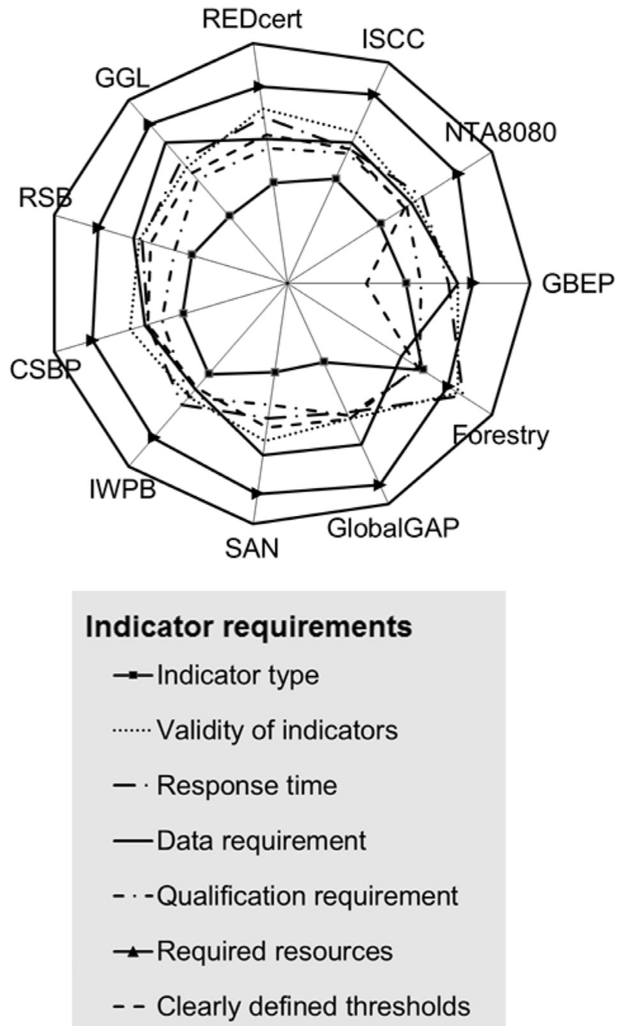

Fig. 2 – Arithmetic mean of the ratings by subcategory of the indicator requirements for the evaluated certification schemes and indicator sets, CSBP [43], GBEP Task Force [14], GGL [44], GlobalGAP [45], ISCC [46], IWPB [47], Netherlands Standardization Institute [48,49], REDcert [50], RSB [51], SAN [52] and forestry, [29,32,53–56]; a detailed explanation of the meaning of each step of the rating scales per indicator requirement is discussed in Section 2.2, and the SEM can be found in Appendix A.

In Fig. 4, the results for the rating of certification scheme indicators are grouped by composite scale to reveal possible further patterns.

3.2.2.1. *Soil quality*. With the exception of the *Data requirement*, soil quality indicators are especially high rated in the forestry indicator set. For the *Data requirement*, the forestry indicator set still performs as well as most of the other certification schemes. The higher rating of the forestry indicator set might reveal some potential for improvement in existing bioenergy certification schemes.

3.2.2.2. *Water quality*. With respect to water quality, most of the certification schemes perform equally well, with the exception of the *Data requirement*. Here, the low rating of the *Indicator type* is very apparent and reflects the dominant use of indicators that assess management practices and not the actual changes in the environmental compartment, i.e., water bodies.

3.2.2.3. *Water availability*. Water availability could be characterized as highly feasible (*Required resources (assessment interval)*, *Qualification requirement*, *Data requirement*) for most of the certification schemes, with the exception of the forestry schemes and GGL, which have low ratings for all of the requirements. This composite scale shows the differences in how well certification schemes chose indicators that optimize the trade-off between requirements, e.g., reliability and feasibility. I.e., a comparable level of reliability and conceptual soundness (*Indicator type, Validity of indicators, Response time*) may be achieved with a high or low resource use (*Required resources (assessment interval)*, *Qualification requirement*, *Data requirement*).

3.2.2.4. *Biodiversity and LU/LUC*. Biodiversity is rated very homogenously by ISCC, REDcert, GGL and IWPB and LU/LUC by REDcert, RSB, SAN, GlobalGAP and forestry indicators. Both groups of certification schemes only use one environmental assessment indicator for biodiversity and for LU/LUC respectively; this indicator is no production of bioenergy feedstocks in areas of high conservational value (ecosystems, species) and no conversion of land use types equivalent to those in the EU RED.

The rather high rating observed, especially for biodiversity, can be explained by the nature of the change because the coupling of biodiversity loss to land-use change facilitates the assessment for most of the requirements. Biodiversity gains higher indicator feasibility and reliability and conceptual soundness from land-use change indicators.

Both the biodiversity and LU/LUC indicators also show the extent to which certification schemes exclusively fulfill and go beyond the underlying legislation. Here, the question is how detailed legislation should define environmental impacts that are to be avoided. Assuming that a large abundance of an indicator in the schemes is equal to the relevance, it can be said that the clear indicator definition by EU RED is suitable. This indicator is also used by other certification schemes than those complying with the EU RED. However, this indicator is most likely not sufficient to comprehensively cover the major environmental impacts if only this legal minimum is assessed by certification schemes. Such clearly defined legislation might even hinder the competition among certification schemes to find an optimal solution for comprehensive detection of environmental impacts.

3.2.2.5. *Other*. The following composite scales are not completely assessed by the respective scheme. These certification schemes lack direct environmental assessment indicators for some of the composite scales: soil quality (GGL), water availability (REDcert) and LU/LUC (GGL, CSBP) (value 0). Indicators that do not belong to any composite, i.e., indicators grouped under *Other*, are largely missing. *Other* indicators only occur in the GBEP, IWPB and forestry schemes, as shown in Fig. 4, and contain only three indicators on sustainable

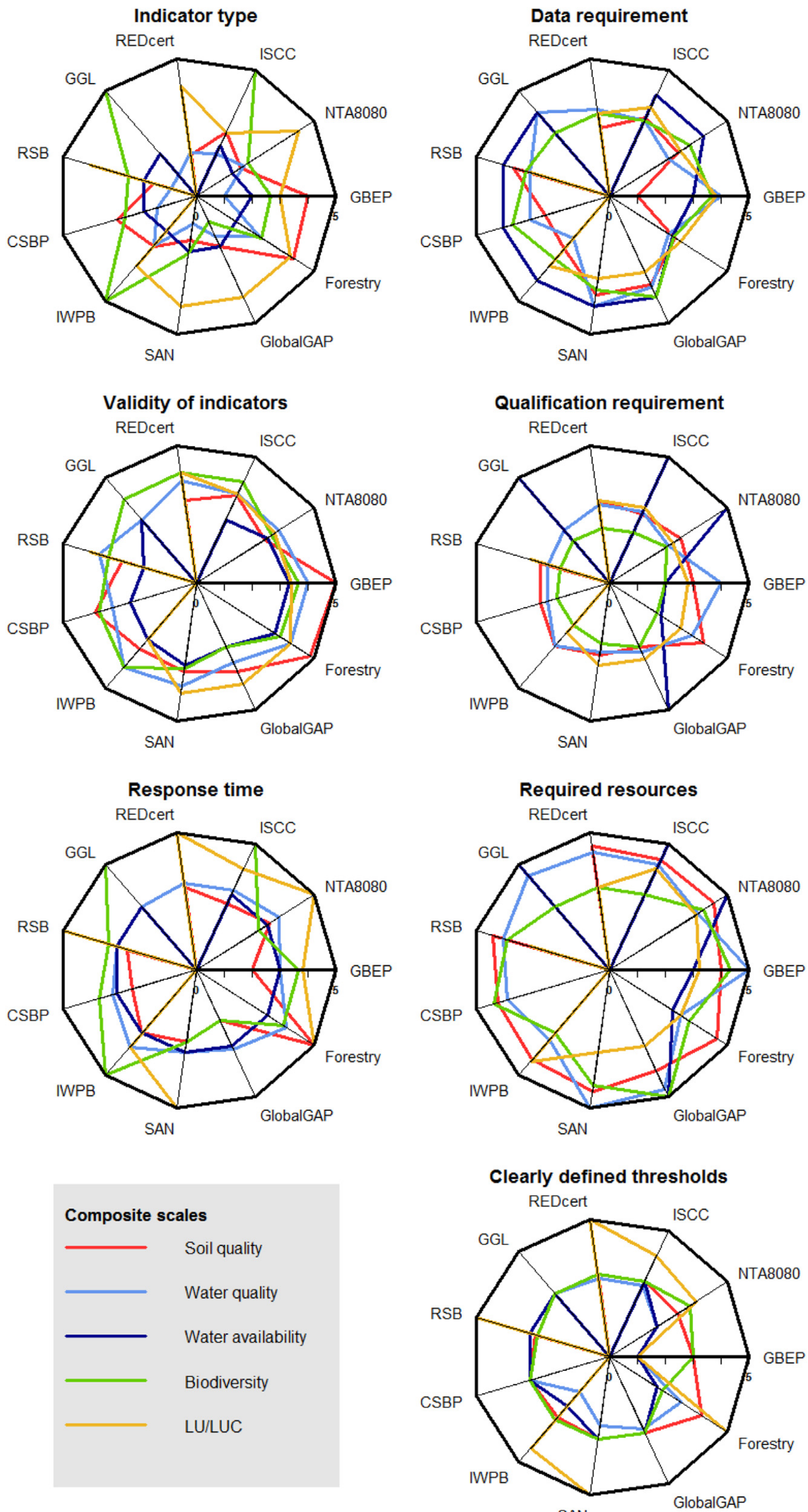

Fig. 3 – Arithmetic mean for each indicator requirement subcategory disaggregated by composite scale and certification scheme/indicator set. Five is the best rating; zero indicates a lack of direct environmental assessment indicators for the composite scale and certification scheme. The SEM can be found in Appendix A.

harvest levels, which are predominately related to forestry. If indicators for the different composite scales are missing for a certification scheme, they are either neglected by the respective certification scheme or the scheme uses no direct environmental impact assessment indicators, as described in Section 3.2.1.

3.3. Comprehensiveness and quality of environmental indicator sets

The certification schemes and indicator sets for bioenergy production are mapped to the ESS cascade as described in Section 2.3 and as displayed in Appendix A.

3.3.1. Comprehensiveness of indicators and causal links for system representation

The comprehensiveness of the system representation in these schemes is shown in Table 4.

Human land use activities can be identified as the most comprehensively covered component of the ESS cascade for most of the schemes reviewed, except for GBEP and ISCC.

This pattern might be explained by the greater feasibility of assessment rather than the relevance of the biophysical processes; see the less comprehensive coverage of ecosystem structures and processes and ESS and the necessity that certification schemes demonstrate sustainability at a local scale instead of the required assessment at a regional scale for other indicators, and see Fig. 1 in Section 2.3. In contrast, the disproportionately small number of indicators to be assessed at a regional scale renders it very likely that certification schemes miss cumulative effects. Cumulative effects are only harmful if a farming practice is applied throughout a region. For example, a crop and the respective fertilizer and pesticide application might only cause significant impacts on water quality if repeatedly applied within a catchment. This problem is addressed by NTA 8080 and IWPB, which both include indicators for off-site impacts, such as the Biological Oxygen Demand. GBEP has a large share of indicators that are beyond the local scale, but this share can very likely be attributed to its difference in purpose. GBEP indicators have been developed for national assessments [14] rather than for certifying single producers.

Ecosystem capacity is considered in most of the certification schemes; however, in RSB ecosystem capacity is not explicitly considered (yellow color) or is not considered (white color), as shown in Appendix A.

An explanation for the lack of thresholds or target values might be the flexibility required to consider the applicability globally and for multiple feedstocks. The indicators need to be equally applicable to different feedstocks that are grown under various environmental conditions and alongside various ecosystems associated with a large variability in ecosystem capacity. Here, clear target values are neither feasible nor practical. However, a methodology for the derivation of the ecosystem capacity can be given. A positive example is the RSB; see Fig. 5. Usually, a threshold is set for the SOC content for several certification schemes. However, the SOC content is only expected to reveal significant changes from changes in management practices, e.g., tillage regime, after a long time lag of at least five to ten years [81]. Because the reviewed certification schemes do not consider such a time lag in their certificate, such a threshold for SOC will be unlikely to have an impact on the certification decision. Only severe changes of the SOC content over the respective time frame might have an impact.

3.3.2. Quality of indicators and causal links for system representation: exemplary cases

The quality of the system representation is analyzed in the examples in Fig. 5; i.e., how certification schemes translate the human–environment interactions and the biophysical cause–effect relationships. As mapped in Fig. 5, the water availability indicators from GGL show that the central aspect of the certification schemes is often driver indicators for management practices, and these indicators should partly consider biophysical processes (2.). These biophysical processes are usually not specified. As an example, indicators are defined as follows: "Data about: climate, water [...] are collected on a regular basis." [44]. In addition, it is required that practices are applied to enhance the use of scarce water resources: "4.1 Efficiency and productivity of agricultural water use for better utilization of limited water resources has to increase" [44]. Neither the practices (3.) nor the ecosystem capacity of a scarce water resource (4.) are defined. Missing indicators and open formulations for indicators often result in imprecisely formulated causal links (5.). In contrast to the previous examples, for GBEP, shown in Appendix A, clearly defined indicators, which result in equally clear causal links, can be found.

A higher accuracy of the defined causal links facilitates environmental performance measurements and the determination of options for improvement. Predictions for the alteration of one parameter allow the direction of the change in another indicator to be determined qualitatively or even quantitatively. For example, excluding land cover types such as peatlands from feedstock production reduces the sustainable yield of a region by the theoretical biomass yield of peatland. As shown in Fig. 5, compared with RSB, a deficiency of both GBEP and GGL is the incomprehensive coverage of most of the components of the ESS cascade.

In contrast to GGL, RSB more comprehensively covers the ESS cascade. Despite the greater comprehensiveness, qualitative deficiencies can be shown for examples of the biodiversity indicators from RSB. Preferably, the indicators used are spatial indicators of biodiversity (1.) and not indicators that directly demonstrate ecosystem functioning, such as species richness and evenness indices, e.g., Shannon index, or the abundance of indicator species (2.). The typically chosen spatial indicators and indicators on conservation practices focus on endangered or protected species and habitats (3.).

Possible explanations for the prevailing indicator choice might be:

a. The requirements of the underlying legislations, i.e., the EU RED, govern the indicator choice.
b. Because of their higher risk of extinction, highly vulnerable species and habitats have greater importance for the public or for nature enthusiasts [82].
c. The availability of data for endangered species and habitats is widely available for many parts of the world. Data on

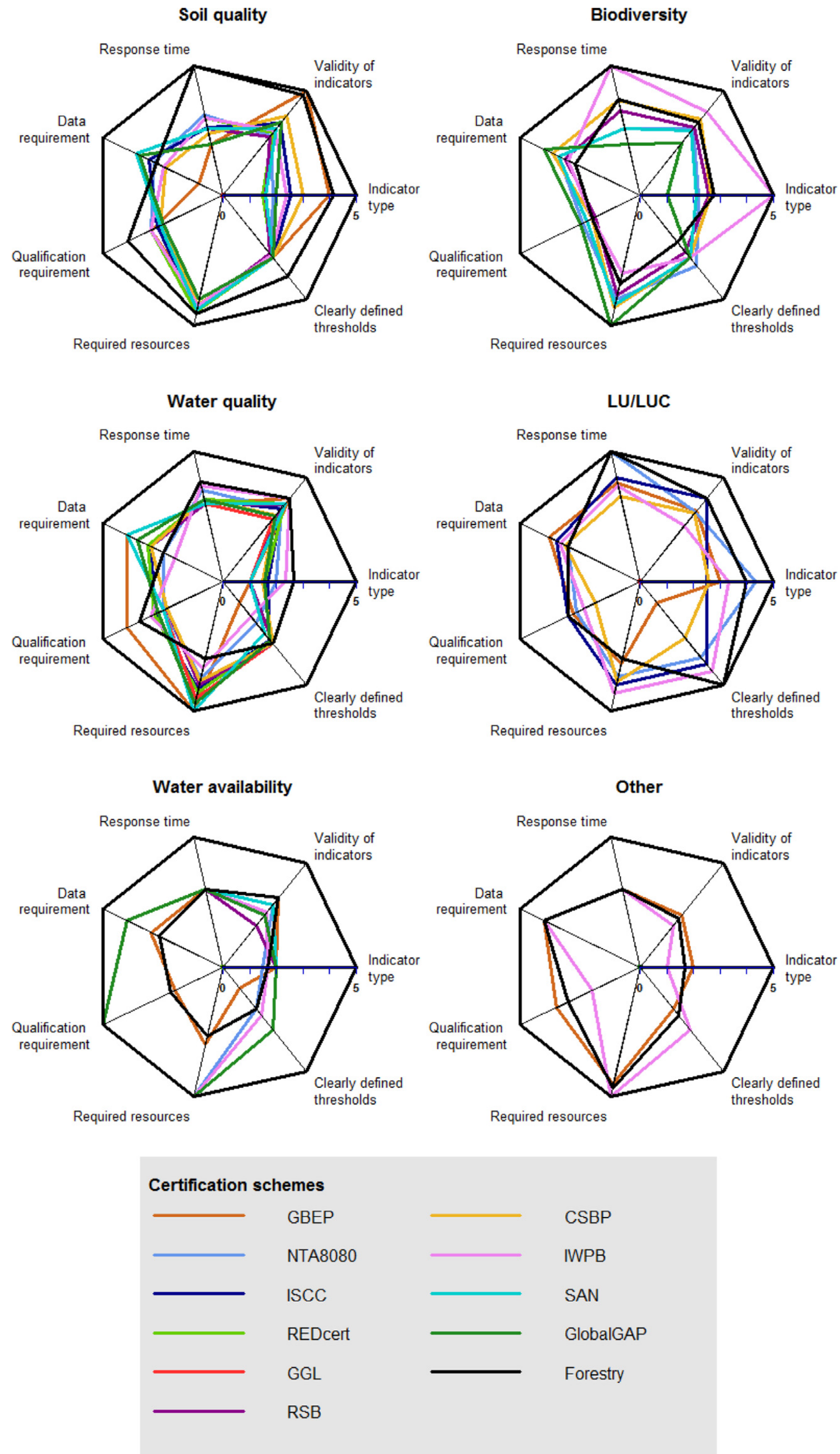

Fig. 4 – Arithmetic mean for each composite scale disaggregated by indicator requirement subcategory and certification scheme/indicator set. Five is the best rating; zero indicates a lack of direct environmental assessment indicators for the indicator requirements and certification scheme. The SEM can be found in Appendix A.

species and habitats of less concern is not collected as extensively [68]. Therefore, data availability seems to be better for indicators on endangered species.

d. Indicators on ecosystem function must be adapted to the local context, i.e., indicator species, other indicanda of ecosystem functioning and species richness greatly vary by both location and ecosystem.

The most common case in which causal links in certification schemes are defined is when management practices are to be applied to minimize the use of ESS and the creation of disservices is respectively compared with an uncertified alternative in feedstock production. This case is revealed for RSB (3.) and GGL in Fig. 5.

Such an approach neglects the underlying ecosystem structures and processes in the indicator definition. Certification schemes assume a shortened causal link from human land use activities to the ESS and ignore the often directly affected ecosystem structures and processes. Currently, certification schemes are unlikely to allow the measurement and comparison of the environmental performance of bioenergy feedstocks. First, certification schemes, as shown for the example in Fig. 5, partially do not cover the obviously affected ESS. For example, biomass (use) is neglected as an indicator although this indicator could easily be determined. Missing indicators are not only those indicators obtained with more effort or technical skills, such as the impact on the minimum and peak flow of surface waters. Secondly, a large proportion of causal links that are represented by the reviewed certification schemes map the interactions between but not within the different components of the ESS cascade. Therefore, it is not possible to determine trade-offs and synergetic interactions between different ecosystem services. Thirdly, feedbacks from the use of ESS on ecosystem structures, processes and capacities are mostly not determined, as shown in the mapped certification schemes. Such less comprehensive coverage of the ESS and the causal links renders it impossible to compare the uses and consequently, the environmental impacts of different feedstocks. This deficiency might be because of the nature of the certification schemes to demonstrate compliance with legislation, such as the EU RED or other non-prescriptive rules. The schemes were not originally developed to assess the environmental performances of different feedstocks. Despite this focus, other ESS affected by biomass use could be theoretically used as a multidimensional unit for normalization to allow comparisons of different pathways for biomass provision; this unit would be comparable to the functional unit, e.g., the biomass, in LCAs for energy use or GHG emissions.

3.4. *Limitations of this approach*

One may argue that there is an assessor bias inherent to both the development and application of the rating scales for the indicator and scheme evaluation. Nevertheless, several measures to reduce and reveal such an assessor bias have been taken:

a. The use of empirically applied and peer-reviewed rating scales for agri-environmental indicator systems;
b. The determination of missing rating scales from the range of weak to strong implementation options for bioenergy certification schemes and existing reviews;
c. Ensuring the transparency of the rating by providing detailed descriptions of each rating scale.

Using the mean to aggregate indicators by composite scale, it was necessary to account for the uncertainty of the mean by the SEM, as shown in Appendix A. There are only a few cases in which the arithmetic mean does not well represent the composite scale. Therefore, the enhanced clarity of the composite scales for each indicator individually should be valued higher. There may be more accurate clustering options than the arithmetic mean, but those options would require complete data sets. Because they do not include indicators for all composite scales, several certification schemes, namely REDcert, GGL, and CSBP, would have had to be excluded. The same problem applies to tests for the internal consistency of the composite scales, such as Cronbach's alpha test, which could not be used because the data sets were incomplete. Because only 3 of 87 indicators could not be grouped to the chosen composite scales, as given by the environmental impact categories, the expert-based approach seems to be sufficient.

Table 4 – Comprehensiveness of system representation in certification schemes and indicator sets; better ratings mean that more indicators are covered for the different components of the ESS cascade (Fig. 1 in Section 2.3), i.e., the representation of the function of the affected ecosystem and the used ESS. For causal links, the certification schemes are compared with their peers. The certification scheme with the highest number of causal links has the best rating and is used as a benchmark.

Certification schemes	Indicators				Causal links
	Ecosystem structures and processes	Ecosystem capacity	Ecosystem services	Human land use activities	
GBEP	–	–	+/–	+/–	–
NTA8080	–	–	+/–	+	+/–
ISCC	–	–	–	+	–
REDcert	–	–	–	+	+/–
GGLS2	+/–	–	+/–	+	+/–
RSB	+/–	+	+/–	+	+
CSBP	–	–	+/–	+/–	+/–
IWPB	–	–	+/–	+	+/–
SAN	+/–	–	+/–	+	+
GlobalGAP	–	–	–	+	–
Forestry	+/–	+/–	+/–	+	+/–

Coverage of indicators: >66.6%: +, 33.4–66.5%: +/–, <33.3%; –.

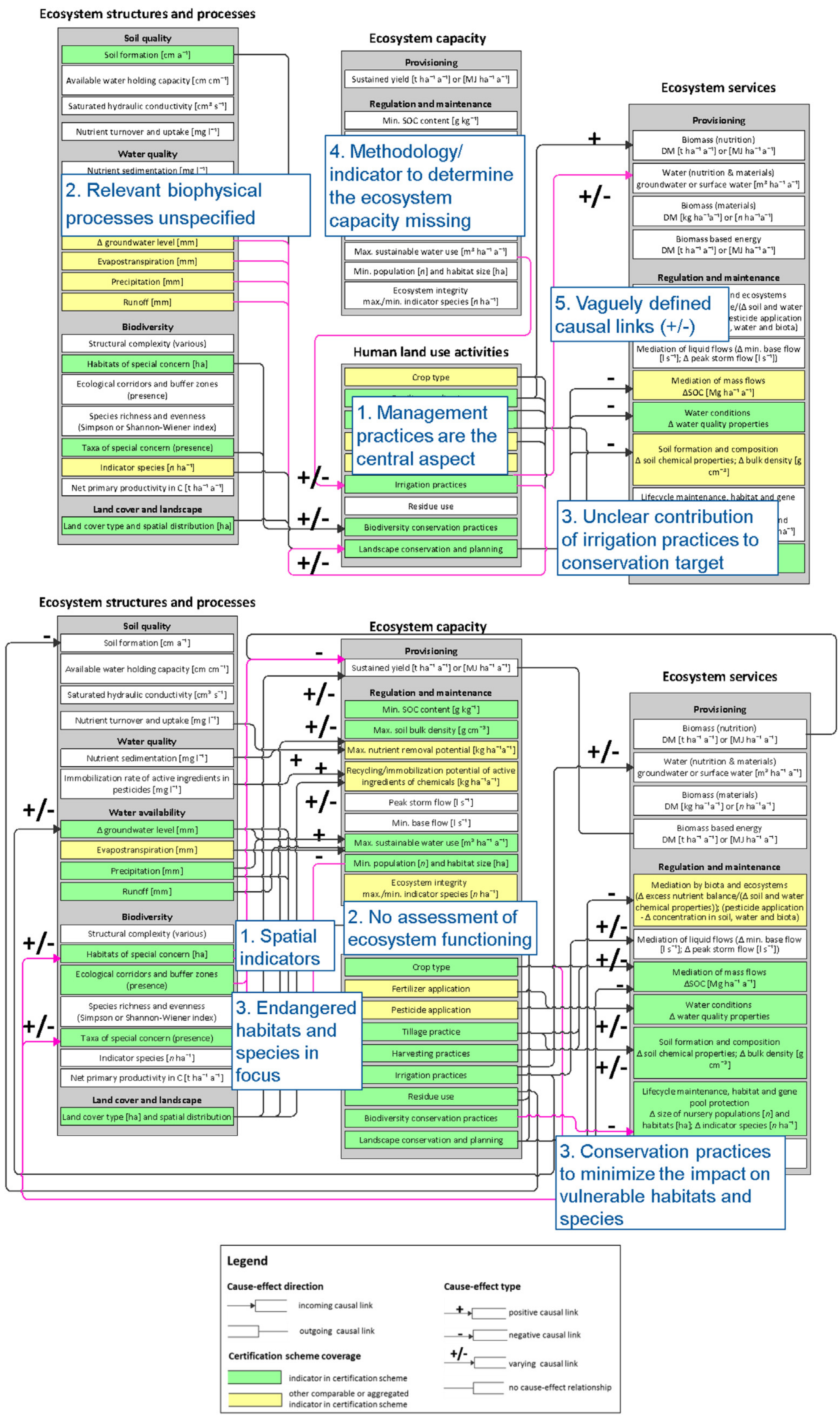

Fig. 5 – ***(upper part)*** **Water availability indicators from GGL mapped onto the ESS cascade.** ***(lower part)*** **Biodiversity indicators from RSB mapped onto the ESS cascade; common characteristics and deficiencies are indicated in the numbered boxes.**

Empirically, the ESS cascade has been used to assess the impact of human appropriation for purely scientific purposes in a number of cases already, e.g., the studies by Kandziora et al. [64], Maes et al. [63], Petz and van Oudenhoven [61]. Such science-focused studies partially may not reflect practical needs. For example, indicators at the catchment scale are not necessarily suitable to certify individual farmers although these indicators are scientifically more appropriate. In addition, the scope of this study on local/regional environmental impacts required the exclusion of global environmental impacts (e.g., air quality). Therefore, a smaller number of interactions with the related ESS, e.g., the atmospheric composition and climate regulation, are missing. Nevertheless, it is unlikely that a few additional ESS would significantly change the relatively clear patterns shown for the included ESS.

3.5. *Results in the context of existing and possible future research*

This section sets the findings of this study in relation to existing research and outlines future research needs.

3.5.1. *Usefulness of precise and harmonized legislation on environmental impacts as baseline for certification schemes*

Biodiversity and LU/LUC, as composite scales, demonstrate that there is a convergence of certification schemes. The results by Van Dam et al. [13] noting the abundance of spatial biodiversity indicators for endangered habitats and species can be confirmed. The actual change in biodiversity is typically not assessed in the evaluated certification schemes, but it is stated to be hardly possible by current schemes and requiring beyond farm scale assessments [12]. For biodiversity, the hypothesis that precise definitions of the underlying legislation such as the EU RED might hinder the use of more reliable impact indicators seems relevant. In particular, other composite scales with less precise definitions, e.g., the Water Framework Directive in the EU, or with no underlying legislation, such as the scale for water quality, show a larger variety of indicators. Such convergence caused by precisely defined legislation indicates that exclusive peer comparison in existing review papers (e.g., Van Dam et al. [26]) does not completely reveal the limitations and potential improvements.

An additional research-based indicator set, such as the analytical framework developed in this study, revealed further limitations and potential improvements. Based on this analytical framework, limitations in the qualitative and quantitative representations of environmental impacts and the use of ESS in certification schemes could be shown. Some certification schemes are good examples for selected aspects of the assessment of environmental sustainability. Improvements may be achieved by combining the comprehensiveness of RSB with the quality of GBEP, for example. The focus on human land use activity indicators and the largely incomplete assessment of other key functional relationships show that the selection of indicators for certification schemes is driven by feasibility rather than by relevance or reliability. With respect to feasibility, Scarlat and Dallemand [12] recommend striving for a further harmonization of certification schemes through a meta-standard approach or through internationally harmonized minimum sustainability requirements. Their approach might contribute to reduced certification costs, increased feasibility or increased international acceptance of bioenergy certification schemes; these effects are comparable to the developments in forestry certification schemes (e.g., FSC and PEFC). However, enhanced reliability and conceptual soundness of certification schemes require empirical tests or comparisons with a research-based indicator set. The converging biodiversity and LU/LUC indicators have shown some limitations of peer comparison for certification schemes and missing improvement options from academia.

3.5.2. *Trade-off between a reliable sustainability assessment and securing feasible compliance with legislation*

The focus on feasibility has been apparent in the indicator evaluation in Section 3.2. Existing studies (e.g., Van Dam et al. [13] or Lewandowski and Faaij [2]) identifying the predominant use of feasible causal indicators can be confirmed. Additionally, recent versions of certification schemes, such as the draft from IWPB issued after the findings of former studies, have not been improved in this respect. In addition, the necessity of linking different spatial assessment scales in a proper consideration of environmental impacts has been identified by Van Dam et al. [13]. Nevertheless, this requirement is still only rarely overcome, e.g., by GBEP. With respect to feasibility, *Data requirement* and *Required resources* could be observed to be drivers for indicator selection. Similarly, the weak inclusion of ecosystem capacities, i.e., thresholds or target values, or the use of causal indicators without thresholds is deficient with respect to both feasibility and conceptual soundness.

3.5.3. *Options to improve current certification schemes*

The interactions (causal links) between and within the different components of the environmental systems mapped to the ESS cascade often seem to be incomplete and/or only weakly specified; this incompleteness makes quantification of the interactions difficult or even impossible. This limitation could be improved after specification of the causal links. Incomplete indicator sets do not favor the reliable (environmental) performance measurement of feedstocks. Bioenergy certification schemes have been developed to demonstrate compliance rather than to measure and compare the environmental performances of different feedstocks, confirming Diaz-Chavez [29]. In addition, only the compliance or non-compliance with the certification scheme is of interest not the variable degrees of under-/over-compliance of different feedstocks and producers under different environmental conditions. Mostly likely, future certification schemes could consider different degrees of compliance, e.g., different threshold levels, since too high requirements for producers with low financial means may hinder them to participate [2]. Implementation options could be an extension to the current differentiation of mandatory and facultative requirements used in several certification schemes, e.g., NTA 8080. This approach might (i.) raise the information content of certification schemes by visualizing different degrees of environmental performance. (ii.) This approach also facilitates access for small shareholders in developing countries if they initially only need to comply with less strict thresholds. (iii.) This approach could also be used as a strong marketing tool.

4. Conclusions

In this study, we evaluated existing indicator sets and certification schemes to assess the environmental sustainability of different feedstocks for bioenergy. No outstanding certification scheme could be identified. Nevertheless, certain available schemes are better than others for assessing the selected environmental impact categories. To date, the proliferation of schemes, which was noted by several authors [12,13,26], has not led to significant changes in the use of reliable and conceptually sound indicators. Instead, schemes strive for feasibility in the indicator choice by complying with existing legislation or consumer expectations. For legislators, potential conclusions could be (i) to require certification schemes and academia to develop more reliable, but still feasible and cost-effective indicator sets, which at least cover the major underlying ecosystem structures and processes, and/or (ii) to consider a methodology to assess the capacity of an ecosystem, i.e., a methodology to determine threshold values for sustainable production. As a second step, certification schemes could assess well-defined causal links and feedbacks for biomass production; for example, schemes could use the adapted versions of the ESS cascade as an analytical framework. The suggested improvements would contribute to increased reliability in the identification of the environmental impacts of bioenergy feedstocks. As an additional benefit, the improved representation of ecosystem functions and feedback mechanisms will facilitate assessments of the interaction between different ESS, such as biomass use, water use or regulating ESS. In further empirical studies, it will be especially interesting to find out, under which conditions cause-related indicators reliably identify sustainable production and for which cases such indicators do not reveal sustainability deficiencies. Beyond the environmental impacts targeted in this study, further social or economic impacts must be considered in bioenergy certification to enable a more comprehensive comparison of alternative feedstocks.

Acknowledgments

This work was partly funded by the by EU FP7 project SECTOR (282826, http://www.sector-project.eu/) (Production of Solid Sustainable Energy Carriers by Means of Torrefaction). The authors thank Ameur Manceur for advice on statistical methods and five anonymous reviewers for their comprehensive and helpful comments.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.biombioe.2014.03.041.

REFERENCES

[1] Demirbas A. Political, economic and environmental impacts of biofuels: a review. Appl Energy 2009;86(Suppl. 1):S108–17.

[2] Lewandowski I, Faaij APC. Steps towards the development of a certification system for sustainable bio-energy trade. Biomass Bioenergy 2006;30(2):83–104.

[3] Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, et al. The path forward for biofuels and biomaterials. Science 2006;311(5760):484–9.

[4] WG III. In: Edenhofer O, Pichs-Madruga R, Sokona Y, Seyboth K, Matschoss P, Kadner S, et al., editors. IPCC special report on renewable energy sources and climate change mitigation. Cambridge and New York: Cambridge University Press; 2012.

[5] World Commission on Environment and Development. Our common future. Oxford: Oxford University Press; 1987.

[6] UN. Johannesburg declaration of sustainable development. World Summit on Sustainable Development. 2002 August 26-September 4. Johannesburg.

[7] Cohen B, Winn MI. Market imperfections, opportunity and sustainable entrepreneurship. J Bus Ventur 2007;22(1):29–49.

[8] Hill J, Nelson E, Tilman D, Polasky S, Tiffany D. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proc Natl Acad Sci U S A 2006;103(30):11206–10.

[9] Robbins M. Policy: fuelling politics. Nature 2011;474(Suppl. 7352):S22–4.

[10] Efroymson RA, Dale VH, Kline KL, McBride AC, Bielicki JM, Smith RL, et al. Environmental indicators of biofuel sustainability: what about context? Environ Manage 2013;51(2):291–306.

[11] McBride AC, Dale VH, Baskaran LM, Downing ME, Eaton LM, Efroymson RA, et al. Indicators to support environmental sustainability of bioenergy systems. Ecol Indic 2011;11(5):1277–89.

[12] Scarlat N, Dallemand JF. Recent developments of biofuels/bioenergy sustainability certification: a global overview. Energy Policy 2011;39(3):1630–46.

[13] Van Dam J, Junginger M, Faaij APC. From the global efforts on certification of bioenergy towards an integrated approach based on sustainable land use planning. Renew Sustain Energy Rev 2010;14(9):2445–72.

[14] GBEP Task Force. The Global Bioenergy Partnership sustainability indicators for bioenergy. Rome: GBEP Secretariat – FAO; 2011.

[15] Von Blottnitz H, Curran MA. A review of assessments conducted on bio-ethanol as a transportation fuel from a net energy, greenhouse gas, and environmental life cycle perspective. J Clean Prod 2007;15(7):607–19.

[16] Haes HAU, Heijungs R, Suh S, Huppes G. Three strategies to overcome the limitations of life-cycle assessment. J Ind Ecol 2004;8(3):19–32.

[17] Koellner T, Geyer R. Global land use impact assessment on biodiversity and ecosystem services in LCA. Int J Life Cycle Assess 2013;18(6):1185–7.

[18] Koellner T, Baan L, Beck T, Brandão M, Civit B, Margni M, et al. UNEP-SETAC guideline on global land use impact assessment on biodiversity and ecosystem services in LCA. Int J Life Cycle Assess 2013;18(6):1188–202.

[19] Milà i Canals L, Rigarlsford G, Sim S. Land use impact assessment of margarine. Int J Life Cycle Assess 2013;18(6):1265–77.

[20] Directive 2013/draft/EU Directive of the European Parliament and of the council on sustainability criteria for solid and gaseous biomass in electricity and/or heating and cooling and biomethane injected in the natural gas network; 2013.

[21] Dale VH, Lowrance R, Mulholland P, Robertson GP. Bioenergy sustainability at the regional scale. Ecol Soc 2010;15(4):23.
[22] Rametsteiner E, Simula M. Forest certification—an instrument to promote sustainable forest management? J Environ Manage 2003;67(1):87–98.
[23] Directive 2009/28/EC. On the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC.2009.04.23. Off J Eur Union 2009;L 140:16–62.
[24] Kaphengst T, Ma MS, Schlegel S. At a tipping point? How the debate on biofuel standards sparks innovative ideas for the general future of standardisation and certification schemes. J Clean Prod 2009;17(Suppl. 1):S99–101.
[25] Junginger M, van Dam J, Zarrilli S, Ali Mohamed F, Marchal D, Faaij A. Opportunities and barriers for international bioenergy trade. Energy Policy 2011;39(4):2028–42.
[26] Van Dam J, Junginger M, Faaij A, Jürgens I, Best G, Fritsche U. Overview of recent developments in sustainable biomass certification. Biomass Bioenergy 2008;32(8):749–80.
[27] Searchinger T, Heimlich R, Houghton RA, Dong F, Elobeid A, Fabiosa J, et al. Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. Science 2008;319(5867):1238–40.
[28] Sikkema R, Steiner M, Junginger M, Hiegl W, Hansen MT, Faaij A. The European wood pellet markets: current status and prospects for 2020. Biofuels Bioprod Bioref 2011;5(3):250–78.
[29] Diaz-Chavez RA. Assessing biofuels: aiming for sustainable development or complying with the market? Energy Policy 2011;39(10):5763–9.
[30] Haines-Young R, Potschin M. The links between biodiversity, ecosystem services and human well-being. In: Raffaelli D, Frid C, editors. Ecosystem ecology: a new synthesis. BES Ecological Reviews Series. C. Cambridge: Cambridge University Press; 2010.
[31] Volpi G. EU policy on bioenergy sustainability. [Presentation on the Internet from 5th of December 2012]. Brussels: European Commission; [cited 2013 January 10]. Available from: http://ec.europa.eu/agriculture/consultations/advisory-groups/forestry-cork/.
[32] Stupak I, Lattimore B, Titus BD, Tattersall Smith C. Criteria and indicators for sustainable forest fuel production and harvesting: a review of current standards for sustainable forest management. Biomass Bioenergy 2011;35(8):3287–308.
[33] Cocchi M, Nikolaisen L, Junginger M, Sheng Goh C, Heinimö J, Bradley D, et al. Global wood pellet industry – market and trade study. Task 40: Sustainable international bioenergy trade. IEA Bioenergy; 2011.
[34] Bockstaller C, Guichard L, Keichinger O, Girardin P, Galan M-B, Gaillard G. Comparison of methods to assess the sustainability of agricultural systems. A review. Agron Sustain Dev 2009;29(1):223–35.
[35] Buchholz T, Luzadis VA, Volk TA. Sustainability criteria for bioenergy systems: results from an expert survey. J Clean Prod 2009;17(Suppl. 1):S86–98.
[36] Niemeijer D, de Groot RS. A conceptual framework for selecting environmental indicator sets. Ecol Indic 2008;8(1):14–25.
[37] Payraudeau S, van der Werf HMG. Environmental impact assessment for a farming region: a review of methods. Agric Ecosyst Environ 2005;107(1):1–19.
[38] Svarstad H, Petersen LK, Rothman D, Siepel H, Wätzold F. Discursive biases of the environmental research framework DPSIR. Land Use Policy 2008;25(1):116–25.
[39] Bockstaller C, Gaillard G, Baumgartner D, Freiermuth Knuchel R, Reinsch M, Brauner R, et al. Betriebliches Umweltmanagement in der Landwirtschaft: Vergleich der Methoden INDIGO, KUL/USL, REPRO und SALCA. Colmar: Grenzüberschreitendes Institut zur rentablen umweltgerechten Landbewirtschaftung; 2006.
[40] Bockstaller C, Girardin P. How to validate environmental indicators. Agric Syst 2003;76(2):639–53.
[41] Binder CR, Feola G, Steinberger JK. Considering the normative, systemic and procedural dimensions in indicator-based sustainability assessments in agriculture. Environ Impact Assess 2010;30(2):71–81.
[42] Coste J, Guillemin F, Pouchot J, Fermanian J. Methodological approaches to shortening composite measurement scales. J Clin Epidemiol 1997;50(3):247–52.
[43] CSBP. Standard for sustainable production of agricultural biomass. [Monograph on the Internet updated 2012]. Council on Sustainable Biomass Production; [cited 2012 December 21]; [about 1 screen]. Available from: http://www.csbp.org/.
[44] GGL. Certification requirements Green Gold Label. [Monograph on the Internet updated 2011]. Zwolle: Green Gold Label Foundation; [cited 2012 December 6]; [about 1 screen]. Available from: http://www.greengoldcertified.org.
[45] GlobalGAP. Integrated farm assurance: all farm base/crops base/combinable crops. [Monographs on the Internet updated 2012]. Cologne: FoodPLUS GmbH; [cited 2012 December 8]; [about 4 screens]. Available from: http://www.globalgap.org/uk_en/.
[46] ISCC. ISCC EU system documents. [Monographs on the Internet updated 2011]. Cologne: ISCC System GmbH; [cited 2012 December 4]; [about 5 screens]. Available from: http://www.iscc-system.org/en/.
[47] IWPB. IWPB sustainability principles/IWPB sustainability criteria and indicators. [Monographs on the Internet updated 2012]. Linkebeek: Laborelec; [cited 2012 December 21]; [about 3 screens]. Available from: www.laborelec.be/ENG/initiative-wood-pellet-buyers-iwpb/.
[48] NTA 8080:2009 NTA 8080 – sustainability criteria for biomass for energy purposes; 2009. pp. 1–41.
[49] NTA 8081:2012-04 NTA 8081 – certification scheme for sustainably produced biomass for energy purposes; 2012. pp. 1–27.
[50] REDcert. REDcert – EU system documents. [Monographs on the Internet updated 2012]. Bonn: REDcert GmbH [cited 2012 December 12]; [about 3 screens]. Available from: http://www.redcert.org.
[51] RSB. RSB sustainability standards – EU RED. [Monographs on the Internet updated 2011]. Chatelaine: Roundtable on Sustainable Biomaterials; [cited 2012 December 8]; [about 5 screens]. Available from: http://rsb.org/sustainability/rsb-sustainability-standards/.
[52] SAN. Farm Standards/Group Standards. [Monographs on the Internet updated 2011]. San José: Sustainable Agriculture Network; [cited 2012 December 6]; [About 2 screens]. Available from: http://sanstandards.org/sitio/.
[53] Lattimore B, Smith C, Titus B, Stupak I, Egnell G. Environmental factors in woodfuel production: opportunities, risks, and criteria and indicators for sustainable practices. Biomass Bioenergy 2009;33(10):1321–42.
[54] Newton AC, Kapos V. Biodiversity indicators in national forest inventories. Unasylva 2002;53(210):56–64.
[55] Schoenholtz SH, Miegroet HV, Burger J. A review of chemical and physical properties as indicators of forest soil quality: challenges and opportunities. For Ecol Manage 2000;138(1):335–56.
[56] Smith GF, Gittings T, Wilson M, French L, Oxbrough A, O'donoghue S, et al. Identifying practical indicators of biodiversity for stand-level management of plantation forests. Biodivers Conserv 2008;17(5):991–1015.
[57] De Groot RS, Alkemade R, Braat L, Hein L, Willemen L. Challenges in integrating the concept of ecosystem services

and values in landscape planning, management and decision making. Ecol Compl 2010;7(3):260–72.
[58] MEA Board. In: Hassan R, Scholes R, Ash N, editors. Ecosystems and human well-being: current state and trends : findings of the condition and trends working group. Washington, DC: Island Press; 2005.
[59] Costanza R, d'Arge R, De Groot R, Farber S, Grasso M, Hannon B, et al. The value of the world's ecosystem services and natural capital. Nature 1997;387(6630):253–60.
[60] Potschin MB, Haines-Young RH. Ecosystem services: exploring a geographical perspective. Prog Phys Geogr 2011;35(5):575–94.
[61] Petz K, van Oudenhoven APE. Modelling land management effect on ecosystem functions and services: a study in the Netherlands. Int J Biodivers Sci Eco Services Manage 2012;8(1–2):135–55.
[62] Van Oudenhoven APE, Petz K, Alkemade R, Hein L, de Groot RS. Framework for systematic indicator selection to assess effects of land management on ecosystem services. Ecol Indic 2012;21:110–22.
[63] Maes J, Teller A, Erhard M, Liquete C, Braat L, Berry P, et al. Mapping and assessment of ecosystems and their services: an analytical framework for ecosystem assessments under action 5 of the EU biodiversity strategy to 2020. Luxemburg: Publications Office of the European Union; 2013.
[64] Kandziora M, Burkhard B, Müller F. Interactions of ecosystem properties, ecosystem integrity and ecosystem service indicators—a theoretical matrix exercise. Ecol Indic 2013;28:54–78.
[65] Ojima D, Moran E, McConnell W, Stafford Smith M, Laumann G, Morais J, et al. In: Young B, editor. Science plan and implementation strategy. Stockholm: IGBP Secretariat; 2005.
[66] Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, et al. Global consequences of land use. Science 2005;309(5734):570–4.
[67] CICES. CICES V 4.3. [Spreadsheet on the Internet updated 2013]. University of Nottingham; Nottingham: [cited 2013 January 15]. Available from: http://cices.eu/.
[68] Dale VH, Beyeler SC. Challenges in the development and use of ecological indicators. Ecol Indic 2001;1(1):3–10.
[69] Kienast F, Bolliger J, Potschin M, de Groot RS, Verburg PH, Heller I, et al. Assessing landscape functions with broad-scale environmental data: insights gained from a prototype development for Europe. Environ Manage 2009;44(6):1099–120.
[70] McElhinny C, Gibbons P, Brack C, Bauhus J. Forest and woodland stand structural complexity: its definition and measurement. For Ecol Manage 2005;218(1):1–24.
[71] Wascher DM. Agri-environmental indicators: for sustainable agriculture in Europe. Tilburg: European Centre for Nature Conservation; 2000.
[72] Sposito GE. Scale dependence and scale invariance in hydrology. Cambridge University Press; 1998.
[73] Turner M, O'Neill R, Gardner R, Milne B. Effects of changing spatial scale on the analysis of landscape pattern. Landsc Ecol 1989;3(3–4):153–62.
[74] João E. How scale affects environmental impact assessment. Environ Impact Assess 2002;22(4):289–310.
[75] João E. A research agenda for data and scale issues in strategic environmental assessment (SEA). Environ Impact Assess 2007;27(5):479–91.
[76] BEFSCI. A compilation of bioenergy sustainability initiatives. [Monographs (each certification scheme) on the Internet updated 2010-11]. Rome: FAO; [cited 2012 December 10]; [about 8 screens]. Available from: http://www.fao.org/bioenergy/foodsecurity/befsci/compilation/en/.
[77] Vissers P, Paz A, Hanekamp E. How to select a biomass certification scheme? In: Lammers E, Op de Coul M, editors. Netherlands Programmes Sustainable Biomass. Utrecht: NL Agency; 2011.
[78] EC. Biofuels – sustainability schemes [Internet]. Brussels: European Commission; [cited 2012 December 20]; [about 5 screens]. Available from: http://ec.europa.eu/energy/renewables/biofuels/sustainability_schemes_en.htm.
[79] FSC. Forest Stewardship Council Certification – ensuring environmental, social and economic benefits [Internet]. Bonn: FSC International Center gemeinnützige Gesellschaft mbH; [cited 2012 December 12]; [about 9 screens]. Available from: https://ic.fsc.org/certification.4.htm.
[80] SFI. Requirements for the SFI 2010-2014 program: standards, rules for label use, procedures and guidance. [Monograph on the Internet updated 2010 Jan]. Washington DC: Sustainable Forestry Initiative Inc.; [cited 2012 December 7]; [about 2 screens]. Available from: www.sfiprogram.org.
[81] West TO, Post WM. Soil organic carbon sequestration rates by tillage and crop rotation. Soil Sci Soc Am J 2002;66(6):1930–46.
[82] Raudsepp-Hearne C, Peterson G, Bennett E. Ecosystem service bundles for analyzing tradeoffs in diverse landscapes. Proc Natl Acad Sci U S A 2010;107(11):5242–7.