



Assessing smallholder sustainable intensification in the Ethiopian highlands

James Hammond^{a,*}, Mark van Wijk^a, Nils Teufel^a, Kindu Mekonnen^b, Peter Thorne^a

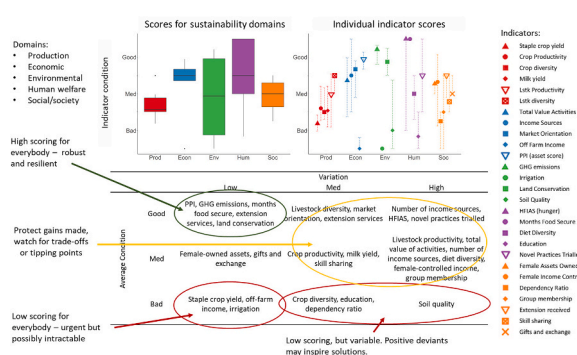
^a International Livestock Research Centre, Nairobi, Kenya

^b International Livestock Research Centre, Addis Ababa, Ethiopia

HIGHLIGHTS

- A five-year program promoting multiple concurrent sustainable intensification technologies was assessed.
- Farm productivity was correlated to improvements in the other sustainability domains, indicating a lack of trade-offs.
- Synergies due to the initially low productivity, the wide range of technologies, and the meaningful farmer participation.
- Crop and livestock productivity was still below attainable levels, and progress was not uniform amongst participants.
- The method presented here is suitable to monitor the transition towards sustainable intensification in portfolio programs.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Dr Laurens Klerkx

Keywords:

Sustainable intensification
Ethiopia
Sustainability assessment
Smallholder
Multivariate visualisation

ABSTRACT

CONTEXT: Sustainable intensification is one approach to increasing food production without undermining sustainability goals. In recent years new tools and indicators have been developed for broad-based assessment of sustainable intensification. However, most of these tools have been applied at field level and assessing individual technologies, while integrated assessments of multiple novel practices at farm-to-village level are lacking.

OBJECTIVE: In this study we develop and apply a data collection, analysis, and interpretation approach that results in a replicable and rapid method for a multi-variate assessment of sustainable intensification.

METHODS: Drawing on a survey of 779 participant farmers, and using the Sustainable Intensification Assessment Framework, we quantified 27 indicators grouped into five domains: agricultural production, economics, environment, human welfare, and social. We applied an expert-led threshold setting exercise to re-scale indicators, permitting aggregated and dis-aggregated visualisation onto a common axis. We developed a graphic evaluation approach to communicate the multiple domain and indicator scores and applied this approach to quantify trade-offs and synergies related to agricultural productivity in four contrasting sites in Ethiopia.

RESULTS AND CONCLUSIONS: In each site there was a notable and significant gradient of production intensity, although average crop and livestock productivity remained well below attainable levels. Higher levels of productivity were correlated with higher scores in the economic, social and agricultural domains, and in some sites they were also positively correlated with the human welfare and environmental domains. In no case was increased production intensity correlated with lower scores in any of the sustainability domains, indicating that in these relatively low-intensity systems increases in productivity will go hand-in-hand with improvements in

* Corresponding author.

E-mail address: j.hammond@cgiar.org (J. Hammond).

<https://doi.org/10.1016/j.agsy.2021.103266>

Received 8 April 2021; Received in revised form 25 August 2021; Accepted 26 August 2021

Available online 8 September 2021

0308-521X/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

most other sustainability domains. Synergies can overrule trade-offs in these smallholder systems in Ethiopia, if managed well.

SIGNIFICANCE: This is one of very few studies of multiple sustainable intensification technologies implemented concurrently at the household to community level. Most studies focus on individual technologies or practices within specific niches of the farming system. The method could be developed further for efficient application to various large-scale development or intensification projects, and could potentially make use of existing smallholder information databases.

1. Introduction

In order to meet the demands of present and future generations, agriculture needs to deliver not only food and fibre, but also needs to deliver a wide range of other social goods, especially in Africa. These include increased supply of nutritious foods; a decent income for farmers regardless of age or gender; contribution to the growth in national and local economies; and ideally environmental improvement, or at least no environmental degradation (Glopan, 2020; HLPE, 2020; IPCC, 2020). If this tall order can be met, agriculture, food supplies, ecosystem services, and the development status of rural communities would all be improved and more resilient than at present.

1.1. The evolution of sustainable intensification

One of the narratives operating within these broad themes is sustainable intensification. Sustainable intensification essentially aims to produce more nutritious food by using the same or less resources. Sustainable intensification is defined by the *objective*, rather than specific technologies, and a plethora of technical and behavioural options are relevant, depending on local context and needs (Pretty, 2008). The paradigm of sustainable intensification came into the mainstream around 2010 (Godfray et al., 2010; Foley et al., 2011; FAO, 2011; Tilman et al., 2011), but the degree to which sustainable intensification has been achieved is still the topic of much investigation (Pretty et al., 2018; Vang Rasmussen et al., 2018; Cassman and Grassini, 2020). Sustainable intensification was originally formulated as relating to environmental issues (Garnett et al., 2013; Godfray, 2015) – such as soil health, water use, greenhouse gas emissions, and the expansion of agricultural land – but others have argued for an expanded-scope definition of sustainable intensification, considering economic and social sustainability goals (Loos et al., 2014; Rockström et al., 2017). A recent review article identified four main thrusts of sustainable intensification research (Cassman and Grassini, 2020): the expanded scope definition; closing yield gaps in lower-income countries; increasing crop diversity and the production of nutrient dense foods; and developing metrics and measurements of sustainable intensification at the field or farm level. The importance of scale and levels in sustainable intensification research has been established (Thomson et al., 2019), and three levels defined: plant to field scale; farm to landscape scale; and national to regional to global scale. A systematic review of 349 papers investigating sustainable intensification (Weltin et al., 2018) identified that the majority focused at the field scale and on individual technologies or practices. The review recommended that more studies focus on higher spatial scales, assess the simultaneous use of multiple technologies, and improve the context-sensitivity of approaches.

1.2. Agricultural sustainability assessments

Dozens of frameworks exist for agricultural sustainability assessments, these have been reviewed elsewhere (Chopin et al., 2021; Inwood et al., 2018; de Olde et al., 2016). Rapid assessments are useful for learning and drawing attention to sustainability outcomes, whereas full sustainability assessments are more useful in designing detailed and specific actions to improve sustainability (Marchand et al., 2014). Rapid assessments make use of farmers' subjective knowledge; data collection

typically takes a few hours per farm, is low cost and therefore applicable over wide areas. The results are presented in a transparent and low complexity format and are relatively easy to understand. Rapid methods may help to achieve the buy-in required for further action. The weaknesses are subjectivity and lower accuracy of results, which can lead to lower levels of trust in the findings (Marchand et al., 2014). Full sustainability assessments may entail hundreds or even thousands of indicators (Pollesch and Dale, 2015); findings may be complex and require extensive calculations and are more suitable for very high-level assessments (Singh et al., 2012). Various authors agree that a balance must be struck between ease of practicability and comprehensiveness (de Olde et al., 2016; Marchand et al., 2014); and that maintaining “meaningfulness” should be paramount (Gan et al., 2017; Asokan et al., 2020). There is general agreement between agricultural sustainability assessments – and sustainability assessments more broadly – regarding the procedure when conducting an assessment. The steps are: (i) definition of study scope, objectives and stakeholders; (ii) indicator selection; (iii) data acquisition; (iv) index creation – which includes weighting, re-scaling, and aggregation; (v) visualisation and communication of findings; (vi) stakeholder engagement (Inwood et al., 2018; Marinus et al., 2018; Musumba et al., 2017; Kanter et al., 2016).

1.3. A sustainable intensification assessment framework

The framework we selected to guide this investigation is the Sustainable Intensification Assessment Framework (SIAF, Musumba et al., 2017), because of the suitable approach taken as well as the support of the USAID's Feed the Future programme. The useful features of SIAF are: the suitability for an expanded-scope definition of sustainable intensification; consideration of multiple spatial scales; flexibility in terms of indicator selection and consideration of different means of data collection; feasibility for rapid application in low-income locations. Applications of this framework to date have focused on the effects of varying options within individual technological niches of a farm system – for example techniques for the delivery of nitrogen to maize crops in Malawi (Snapp et al., 2018), and the introduction of forage chopping machines to Tanzanian farms (Fischer et al., 2018). In this study, we focus our analysis on a more aggregated level, looking at the outcomes of uncontrolled combinations of novel farm technologies (selected by farmers themselves), and consider these outcomes at the household to village scale. The primary approach to visually communicate findings from the SIAF and other frameworks has been the spider diagram, emphasising the independence of the included domains (e.g. Dominguez-Hernandez et al., 2018). However, here we demonstrate the benefits of an alternative, using box-and-whisker and point plots. We also further develop the efficiency of the approach by collecting the data required for indicator calculations using the well-established RHoMIS (Rural Household Multiple Indicator Survey) system, which includes a standardized household survey tool (e.g. Hammond et al., 2017; van Wijk et al., 2020) as well as associated processes for generating productivity and livelihood indicators. A key objective of this study was to develop an integrated data collection and data analysis approach that results in a replicable and rapid method, allowing such multi-indicator assessments of sustainable intensification to be implemented widely in the developing world. To this aim, we:

- Apply the SIAF framework within the Africa RISING Ethiopia program using data collected with the RHoMIS tool
- Develop a graphic presentation of indicator scores across multiple dimensions for the efficient evaluation and interpretation of the results obtained
- Apply this approach to determine the trade-offs and synergies of different domain and indicator scores related to variations in agricultural productivity
- Evaluate overall the methods used as a generalisable approach to rapid multivariate evaluation of sustainable intensification

2. Materials and methods

2.1. Description of the Africa RISING program

The stated goal of the Africa RISING program is to provide pathways out of hunger and poverty for smallholder farm families through sustainably intensified farming systems that sufficiently improve food, nutrition, and income security, particularly for women and children, and conserve or enhance the natural resource base (Africa RISING, 2019). The program was initiated in late 2011, continues until at least 2021, and is funded by the USAID's Feed the Future Initiative. It operates in three African regions. In this manuscript we refer only to implementation of the program in the Ethiopian highlands. The program in Ethiopia operated in four woredas (districts) of the four main highland regions: Basona Worena in the Amhara region; Endamehonei in Tigray, Lemo in the Southern Nations, Nationalities, and People's Region (SNNPR), and Sinana in Oromia. In each woreda, the program engaged deeply in two kebeles (villages), which are shown in Fig. 1. The findings of this study are representative of project participants within these kebeles; however to facilitate recognition of the study areas, the woreda names have been used.

A wide range of technologies were promoted by the program from 2014 onwards, to all interested farmers living within the research sites. Participants were enrolled in the program via existing networks, namely the kebele (village) administrative offices and department of agriculture offices. Efforts were made to include farmers from the full range of economic and demographic background present within the kebeles, but the risk of unrepresented and excluded groups remains, either because they were not contacted or chose not to participate. Participant farmers were free to select whichever technologies they were interested in and received support in the form of training and materials. About one third of farmers adopting the innovations trialled only one technology, and two-thirds trialled two or more technologies. Around 10% of farmers trialled five or more technologies. The continuation rates were high – on average, 80% of the technologies trialled were still in use by 2018. There were four main themes amongst the promoted technical interventions: livestock feeding, crop production, soil and water management, and mechanisation (Lunt et al., 2018). The livestock feeding technologies were: cultivated forage crops and crop mixtures (oat-vetch, desho grass, phalaris grass, alfalfa, fodder beet, sweet lupin, napier grass, faba bean intercropping); fodder trees (tree lucerne); improved feeding troughs; improved feed storage facilities; and sheep fattening linked to markets. The crop production technologies were: participatory selection of improved varieties (bread wheat, faba bean, potato, durum wheat, malt barley, food barley, chickpea, enset, lentil); tree crops (apple, avocado); and seed production systems (potato seed production and storage, wheat seed production, faba bean seed production). The soil and water management technologies were: soil fertility (manure trials, fertiliser trials, soil testing); soil movement (raised beds, ridge and furrow planting); ponds; water pumps (rope and washer pump, solar pump). The mechanisation technologies were tractor seeding, tractor transport, and tractor-powered pumping of water. Only a small number of tractors were available.

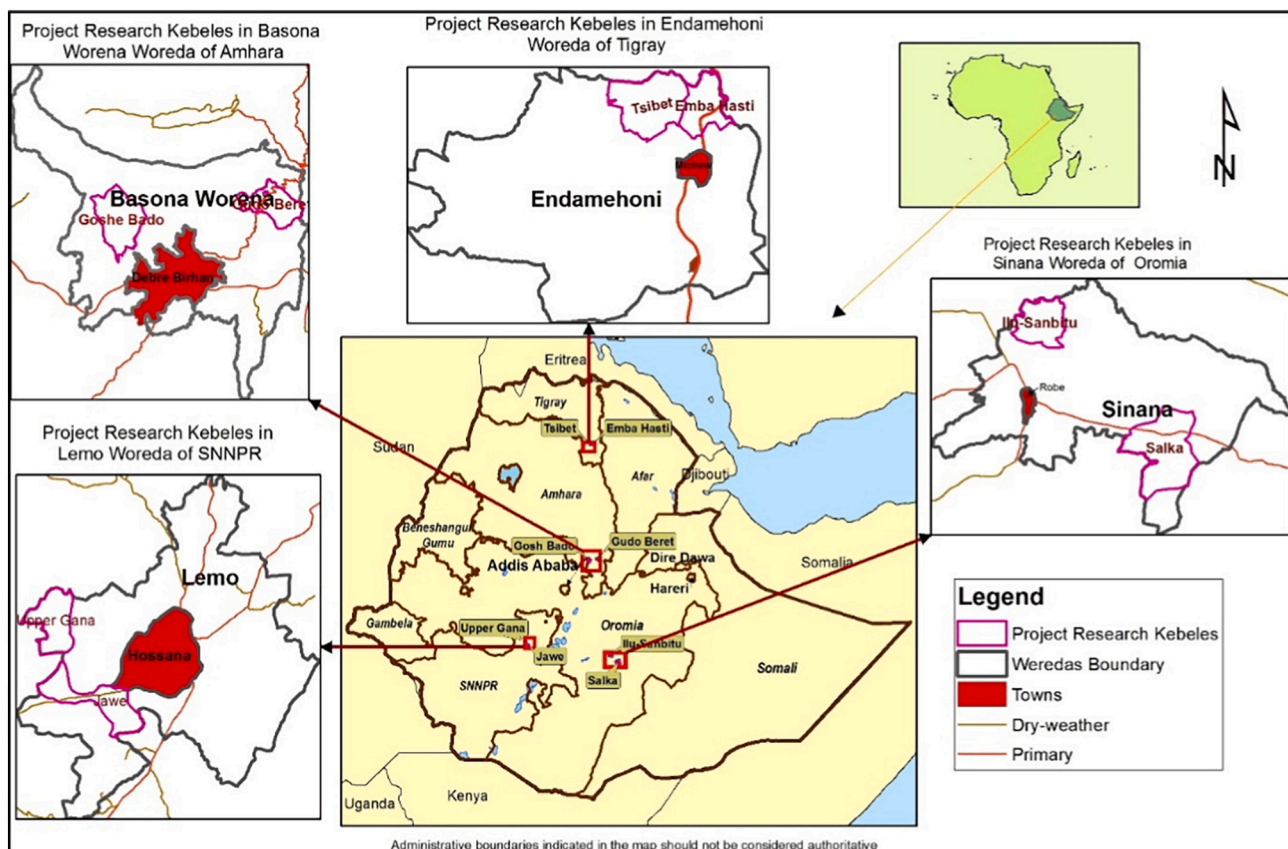


Fig. 1. Map showing the locations of the four study sites in the Africa RISING program, Ethiopia.

2.2. Site descriptions

The four study sites differed considerably in terms of climate and biophysical environment, the agricultural system configurations, and local culture, religion, and demographics (see Table 1). Due to these differences the findings are more meaningful when understood within each study site, rather than comparing between sites.

The Basona Worena site in Amhara was characterised by degraded soils and watersheds with low crop and livestock productivity, and low feed availability, and mixed crop-livestock farming (ILRI, 2015a). The elevation range was 2000–3800 m above sea level (masl), mean annual rainfall of between 900 and 2000 mm per year, mean minimum temperature of 6 °C and mean maximum temperature of 22 °C. The primary crops were wheat, faba bean, barley, potato, with mean farm size 1.3 ha. The main livestock species were cattle, donkeys, sheep, and chicken, and households on average owned 1.9 TLU of livestock (Tropical Livestock Unit, where 1 unit is the mass equivalent of 1 adult cow; Njuki et al., 2011). In a participatory assessment at the start of the project (Ellis-Jones et al., 2013) 33 relevant institutions were identified within the research kebeles. The most important were the kebele administrative offices, which facilitated almost all activities; the kebele agricultural office, which delivered agricultural training and extension, and agricultural cooperatives which delivered inputs. Also important were credit and savings associations and small religious NGOs delivering humanitarian support. The two research kebeles (villages) were Gudo Beret and Goshe Bado.

The Endamehoni site in Tigray was characterised by water scarcity, degraded soils, low crop yields, a shortage of protein-rich fodder, few opportunities for income diversification, and small farm sizes (ILRI, 2015b). Mixed crop-livestock farming was practiced, although with an emphasis on crops. The elevation range is 2800–3000 masl, mean annual rainfall between 600 and 800 mm per year, mean minimum temperature of 2 °C and mean maximum temperature of 15 °C. The primary crops were wheat, barley, faba bean, potato, with mean farm size 0.8 ha. The main livestock species were cattle, sheep, chicken, and donkeys, and households on average owned 1.3 TLU of livestock. Thirty-eight relevant institutions were identified within the research kebeles (Ellis-Jones et al., 2013). The most important were the kebele administrative offices; the department of agriculture extension officers and the farmer training center, and agricultural cooperatives which delivered inputs. Also important were credit and savings associations, religious institutions, and small religious NGOs delivering humanitarian support. The two research kebeles were Emba-Hazti and Tsibet.

The Lemo site in SNNPR was characterised by population pressure, land fragmentation, out-migration, feed shortage, soil acidity, high crop diversity with trees, fruits, horticulture, and field crops, and enset disease (ILRI, 2015c). Mixed crop-livestock farming was practiced, although with an emphasis on crops and fruit. The elevation range is 2000–2500 masl, mean annual rainfall between 900 and 1400 mm per year, mean minimum temperature of 18 °C and mean maximum temperature of 23 °C. The primary crops were wheat, faba bean, teff, and enset,

with mean farm size 1.3 ha. The main livestock species were cattle, donkeys, chicken, and goats, and households on average owned 0.7 TLU of livestock. Eighteen relevant institutions were identified within the research kebeles (Ellis-Jones et al., 2013). The most important were the department of agriculture extension officers and the farmer training center, agricultural cooperatives which delivered inputs, the grain mill, and religious institutions. Also important were the kebele administrative offices, credit and savings associations, and small religious NGOs delivering humanitarian support. The two research kebeles were Jawe and Upper-Ghana.

The Sinana site in Oromia was characterised by productive and often mechanised cereal mono-cropping, larger farms, crop diseases, poor livestock nutrition, and poor human nutrition. Livestock were kept although the main focus was on cereal mono-crops (ILRI, 2015d). The elevation range is 2000–2800 masl, mean annual rainfall between 900 and 1400 mm per year, mean minimum temperature of 6 °C and mean maximum temperature of 20 °C. The primary crops were wheat, emmer wheat, and faba bean, with mean farm size of 3.6 ha. The main livestock species were cattle, horses, donkeys, sheep, and chicken, and households on average owned 1.4 TLU of livestock. Twenty-four relevant institutions were identified within the research kebeles (Ellis-Jones et al., 2013). The most important were the department of agriculture extension offices and farmer training center, agricultural cooperatives which delivered inputs, and animal health clinic. Also important were credit and savings associations, the church, and small religious NGOs. The two research kebeles were Salka and Ilu-Sanbitu.

2.3. Survey description and sampling

The data reported in this manuscript were collected via a survey of participant farmers during April and May 2018, carried out by trained Ethiopian enumerators. The questionnaire was a locally adapted version of the Rural Household Multi-Indicator Survey (RHoMIS, Hammond et al., 2017; van Wijk et al., 2020), which is built on Open Data Kit (ODK) technology (Hartung et al., 2010). The questionnaire covered household demographics; crop and livestock production, management, consumption, and sales; land and natural resource management; off-farm incomes; food security indicators; and gendered control of income and foodstuffs. Unless otherwise specified, all questions related to the entire farm-household and the previous 12 months form the date of the survey. The vast majority of interviews took between 45 and 60 min to complete. Of all respondents, 86% were male and 92% considered themselves to be the head of the household.

The sample was randomly drawn from the complete list of farmer participants in each of the kebeles. The sample size was calculated to allow significant differences of 20% of the mean to be identified within sites, comparing any two groups within the woreda. The sample size calculations were based on program records, population size of the kebeles (with a finite population correction applied), and the mean and variance of indicators of interest gathered in similar studies using the same survey tool. Random sampling was then performed from the

Table 1
Contextual information for the four study sites.

Woreda	Region	Altitude	Climate	Market infrastructure	Farming
Basona Worena	Amhara	2000–3800 m	900–2000 mm rain 6–22 °C.	Good roads and market access.	Mixed, land degraded, watershed management, feed scarce, cattle and dairy.
Endamehoni	Tigray	2800–3000 m	600–800 mm rain 2–15 °C	Markets quite accessible.	Mixed, with crop focus, small farms. Dry with water and feed scarcity.
Lemo	SNNPR	2100–2400 m	900–1400 mm rain 18–23 °C	Good roads and markets, but other infrastructure poor.	Population pressure and outmigration; fruit orchards, crops and livestock.
Sinana	Oromia	2000–2500 m	750–1000 mm rain 6–22 °C	Poor roads, but sales for grains were well organised.	Cereal mono-cropping, larger farms.

household lists to meet the required sample size. In total, 779 interviews were conducted: 148 in Basana Worena, 213 in Endamehonei, 254 in Lemo, and 164 in Sinana.

2.4. Sustainable intensification assessment procedure

2.4.1. Indicator selection

We followed the approach described in the Sustainable Intensification Assessment Framework (SIAF, [Musumba et al., 2017](#)), whereby the selection of indicators should be in relation to the expected outcomes of the innovation under scrutiny. Here we assessed not a single innovation but the net outcome of many technical innovations and the enabling environment established by the Africa RISING program. Accordingly, the selection of indicators was pitched at the household-to-community level, which is where the program outcomes were expected. The SIAF defines five sustainability domains: agricultural production; the economic; the environmental; human welfare; and the social. A long list of indicators are proposed for each domain, with associated methods for measurement and the scale at which the findings are meaningful. This, and the assessment of [Smith et al. \(2017\)](#), guided the selection of indicators for our application of the framework. A further parameter in selection was that the data requirements for each indicator should be achievable through rapid and robust collection by household survey. The indicators selected are listed, and where necessary, explained below:

(i) Agricultural production

The intensity of agricultural production was measured with 6 indicators. Crop productivity is assessed with the yield of the primary staple crop, measured in tons per hectare per year ($t\ ha^{-1}$); and total farm crop productivity, measured in terms of energy content (kilo-calories per hectare per year, $kCal\ ha^{-1}$). In this case, the primary staple crop was either wheat, barley, fava bean, teff, or maize depending on which was produced in the highest quantity on an individual farm. Through re-scaling the crop yields achieved against potential yields (see section 3.4.2) it was possible to compare the yields of differing staple crops. Livestock productivity was assessed with two indicators: the yield of milk per cow, measured in litres per day; and the total livestock productivity, measured in terms of energy produced per unit of livestock owned ($kCal/TLU$). This included the calorie equivalent value of all livestock products (meat, milk, eggs, and wholesale of animals) which were either sold or consumed by the household within one year previous to the survey. Agricultural diversity was assessed using the count of plant species cultivated and the count of livestock species kept.

(ii) Economic

Sustainable economic success was assessed with 5 indicators. The cash value of farm produce at local market conditions, plus any non-farm income, was represented by the “total value of activities” indicator, measured in USD PPP (US dollars adjusted to 2015 purchasing parity power, [World Bank, 2021](#)) per household member per day. The number of independent sources of cash income provided an indication of income resilience. Off-farm income, measured in USD PPP per household per year, was used to indicate income resilience and ability to invest. Market orientation was used to indicate market connectivity and was measured as the percentage of farm produce being sold out of the total value of farm produce. The poverty probability index (PPI) for Ethiopia was used, which provides estimate of the likelihood (0–100%) that a household was below the 1.90 USD per day poverty line, based on observable traits such as asset ownership ([Schreiner and Chen, 2009](#)).

(iii) Environment

Environmental impact of farming activities was represented by 4

indicators. The total warming potential of greenhouse gas emissions per household member per year were expressed as tons of carbon dioxide equivalent (tCO_{2eq}), calculated from the survey data using Tier 1 protocols ([IPCC, 2006](#)). The number of land conservation practices applied per household was used as an indicator of good natural resource management. Practices included incorporation of manure and crop residues in soil, the intentional use of tree or legume species for soil and water benefits, and various physical measures to control water flow and erosion, such as terracing, check dams or bunds. Soil quality was assessed with three basic questions regarding the farmers’ perception of soil fertility, erosion and water holding capacity of their land. “Months per year of irrigation use” was used to indicate water consumption.

(iv) Human welfare

Human welfare was conceptualised as relating to the individual level, as opposed to the higher-level social domain, and was determined through 5 indicators. Food security is a major component of human welfare and was assessed using the household food insecurity of access scale (HFIAS, [Coates et al., 2007](#)) which indicates experience of hunger, the household dietary diversity score (HDDS, adapted from [FAO and FHI360, 2016](#); see [Fraval et al., 2019](#)) which indicates the nutritional quality of the diet, and the number of months during which food access was perceived to be difficult. Innovation capacity was assessed by the education level of the household head, and by the number of novel farming practices the household had trialled during the previous four years.

(v) Social

Within the social domain, 7 indicators were considered. Gender equity was assessed using two indicators: first the number of household assets over which female control was reported, where the assets were land, livestock, crop planning, and access to credit. Second the proportion of the total value of household activities over which females exerted decision-making power (see [Tavener et al., 2019](#)). The conditions for knowledge exchange and information flow were assessed by the number of support groups households participated in; an aggregation of the frequency and quality of extension support received by the household; and the number of peers to whom respondents had shared information about novel agricultural practices. Social reciprocity was assessed by a count of notable gifts and exchanges given and received by households. Labour availability was assessed using the dependency ratio, defined as the ratio of working age adults to non-working age adults and children (above 65 years of age or below 14).

2.4.2. Index creation

In order to permit the meaningful comparison of indicators and aggregation into domains, we applied a novel method of re-scaling the indicator scores. For each indicator, threshold values were selected by which the score could be judged to be “good”, “moderate”, or “bad”. Maximum and minimum scores were also defined for each indicator and used to identify outliers. Credible outlier values were re-scored to the maximum or minimum, and non-credible outliers were excluded. The indicator scores were then re-scaled to a new value between -1 and $+1$, where -1 represented the minimum, $+1$ represented the maximum, -0.33 represented the lower threshold (below which was “bad”) and $+0.33$ represented the upper threshold (above which was “good”). Between each threshold, the scores were re-scaled in a linear manner, but when judged on the complete scale of -1 to $+1$ the scaling was not linear. To illustrate: if an indicator score below 5 out of 10 was considered “bad”, and a score above 8 out of 10 was considered “good”, then the original values between 0 and 5 would be re-scaled between -1 and -0.33 , values above 5 and below 8 would be re-scaled between -0.33 and $+0.33$, and values of 8 or above would be re-scaled between $+0.33$ and $+1$. The selection of reference values is therefore key in

determining the results (as is true for other assessments methods – see Gan et al., 2017 or Acosta-Alba and Van der Werf, 2011). We address this by clearly expressing the thresholds chosen and rational for each (Table 2), and we also present the original indicator scores in Table 6. The re-scaled scores are used for visual representation and aggregation to domain level. The use of subjective language such as “good” and “bad” is intentional in that it conveys the fact that subjective judgements were made, inviting the reader to assess whether they agree with the criteria for that judgement and creating a more meaningful framework for interpreting indicator scores which have been normalised to a unitless common scale (Ebert and Welsch, 2004).

The threshold values identified were normative where possible; otherwise relative values were used (Acosta-Alba and Van der Werf, 2011). Normative reference values are non-site specific and established in scientific or policy literature – examples in this case are the crop yield thresholds for Ethiopian conditions, taken from various scientific literature sources (Cochrane and Bekele, 2018; Lee, 2018; Minten et al., 2018; van Loon et al., 2018), milk yield thresholds (Yilma et al., 2011), or references to international poverty lines. Where values were not available in the literature, we consulted Ethiopian scientists working in agricultural development and familiar with the study sites to suggest appropriate threshold values. Where this was not possible, relative reference values were constructed according to terciles of the study population indicator scores. For indicators where relative threshold values were used, the correct interpretation is that households (or sites) scored “better” or “worse” compared to others in the study, not that the scores are “good” or “bad” within a broader context. However, due to the need to present a large number of indicators visually in a compact

manner, we have used the terms “good” and “bad” for all indicator results as these apply to the majority of the indicators.

Each indicator was calculated for each household, and then aggregated up to groupings for analysis: either terciles of agricultural productivity, or site-level groupings. The median and the inter-quartile range was taken per grouping. Domain scores were then calculated by determining the median and inter-quartile ranges of the of the indicator averages (a so-called “multi-level aggregation” Pollesch and Dale, 2015). In this approach, indicators were weighted equally, and compensation between indicators was permitted (a weak sustainability approach, see Pearce et al., 1994). The presentation of both the aggregated and disaggregated values, presentation of the variation, and transparency in the subjective decisions of normalisation (re-scaling) all follow the recommendations of best practices (Inwood et al., 2018; Gan et al., 2017; Bezlepina et al., 2011; Singh et al., 2012).

2.4.3. Quantifying agricultural intensity

Definitions of intensity usually relate outputs to inputs, but exactly what constitutes outputs and inputs, and how to measure them, differs widely (Smith et al., 2017). Here we quantify agricultural productivity as the key variable to be evaluated, so an output-input-ratio of the agricultural system was used rather than a series of input variables (i.e. production per unit of land and per unit of livestock). We assessed relationships between the intensity variable and other indicators in the SIAF framework, and we also tested whether the uptake of interventions promoted by the Africa RISING program showed any relationship with farm intensity. As such this study is therefore not meant to be a detailed evaluation of the effectiveness of the Africa RISING program, but rather

Table 2

The Indicators selected per Sustainability Domain, the units of measurement, and the threshold values to determine “bad”, “medium”, or “good” scores for each indicator.

Domain	Indicator	Unit	Threshold 1: less than is “Bad”	Threshold 2: greater than is “Good”	Method for setting of thresholds
Production domain	Staple crop yield (wheat/barley/fava bean/teff/maize)	t ha ⁻¹ yr ⁻¹	2.3/2.3/1.0/0.9/3.4	4.7/4.7/1.9/1.8/7.0	One or two thirds of attainable maximum
	Crop productivity	kCal ha ⁻¹ yr ⁻¹	3E+06	6E+06	Population terciles
	Crop diversity	Species grown	5	8	Population terciles
	Milk yield (local breeds/improved/both)	Litres cow ⁻¹ day ⁻¹	0.9/6.3/3.1	1.8/12.7/6.4	One or two thirds of attainable maximum
	Lstk productivity	kCal TLU ⁻¹ yr ⁻¹	7E+04	5E+05	Population terciles
	Lstk diversity	Species kept	2	4	Population terciles
	Total value of activities	USD person ⁻¹ day ⁻¹	1	1.9	Intern'l poverty lines
	Income sources	Count	3	5	Population terciles
Economic domain	Market orientation	% produce sold	<10 or >90	<25 or >75	Moderate market engagement preferable
	Off farm income	USD household ⁻¹ yr ⁻¹	500	1000	Judgement
	PPI (asset-based poverty estimate)	% chance above poverty line	50	75	Judgement
Environment domain	GHG emissions	tCO _{2eq} pers. ⁻¹ yr ⁻¹	>5000	>2000	Compared to international averages emissions
	Irrigation	Months yr ⁻¹	1	>2 & <8	Judgement
	Land conservation	Count of practices	1	4	Judgement
	Soil quality	Farmer perception (0–3)	2	3	Judgement
	HFIAS (hunger)	Severe to no experience of hunger (1–4)	3	4	Indicator guidelines
Human domain	Months food secure	Months yr ⁻¹	10	12	Judgement
	Household dietary diversity	Food groups consumed weekly (0–10)	5	6	Indicator guidelines
	Education	Ordinal (1–7)	4	6	Primary/secondary education
	Novel practices trialled	Count over last 5 years	1	3	Population terciles
	Female assets owned	assets owned or co-owned by females (0–4)	2	3	Judgement
Social domain	Female control	% of livelihood controlled by females	<20 or >90	<35 or >65	Equitable decision making preferable
	Dependency ratio	Ratio of non-workers:workers	<45 or >150	<60 or >100	Judgement
	Group membership	Groups joined yr ⁻¹	1	2	Judgement
	Extension services	Extension services frequency and quality (1–18)	4	9	Judgement
	Skill sharing	Count of people receiving skills	3	10	Judgement
	Gifts and exchange	Count of substantial items given or received yr ⁻¹	<1 or >5	>3	Moderate sharing is preferable

an evaluation of the effectiveness of the SIAP framework to quantify indicator and domain value changes (and their potential trade-offs and synergies) along a production intensity gradient, as presented by the Africa RISING program. For the quantification of agricultural productivity within the mixed crop-livestock systems present in the four sites, we used a combination of crop and livestock productivity, expressed in kCal produced per unit land area and per unit of livestock holdings. We summed the two measures to give an overall intensity score by which to rank the households. Comparison between intensity terciles were made using the Wilcoxon-rank signed sum test, and correlation between intensity and sustainability indicator or domain scores was assessed using Spearman's rank correlation coefficient. Before assessing the correlation of crop and livestock productivity with sustainability indicators in the agricultural production domain, we removed these from the domain scores. However, we acknowledge that some level of auto-correlation between the production domain indicators and our measure of productivity is likely.

2.4.4. Communicating sustainability

Many sustainability assessments visualise findings using spider diagrams (also known as radar charts or amoeba charts), as they allow concurrent visualisation of a large number of indicators and comparison of multiple scenarios (Kanter et al., 2016). However, spider diagrams have also been criticised because the circular layout hampers intuitive

comparison between indicators and they do not permit illustration of uncertainty (Miettinen, 2014). Also common are radial diagrams and petal diagrams which are prone to some of the same critiques. Spatially explicit maps (heat maps) show promise for communication of sustainability assessments, but require a spatially explicit sampling strategy and interpretation (Goldstein et al., 2012). We selected box-and-whisker plots to show aggregated domain scores and point plots to show indicator scores, both with error bars to show the variation in scores. The general layout of the analyses follows the setup shown in Fig. 2. The purpose of showing both the aggregated domain scores and the indicator scores is to facilitate comprehension by the reader – the domain scores give a quick but superficial understanding of the situation, and provide a “way in” to the more complex but deeper description provided by the indicator scores.

The analysis results show the scores of each sustainable intensification domain (coloured boxplots) as well as the median and inter-quartile range of each indicator (the point plots using coloured symbols). Two measures were used: i) the average value of domain and indicator scores to summarise which indicators scores are considered to be in “good”, “medium”, or “bad” condition for a site, and ii) the degree of variation observed, as an indication of the potential of a domain and indicator score to be improved within current site conditions. These populate the prioritisation framework in the lower panel of the figure. Figures were made using the package ggplot2 (Wickham, 2016) in the R software

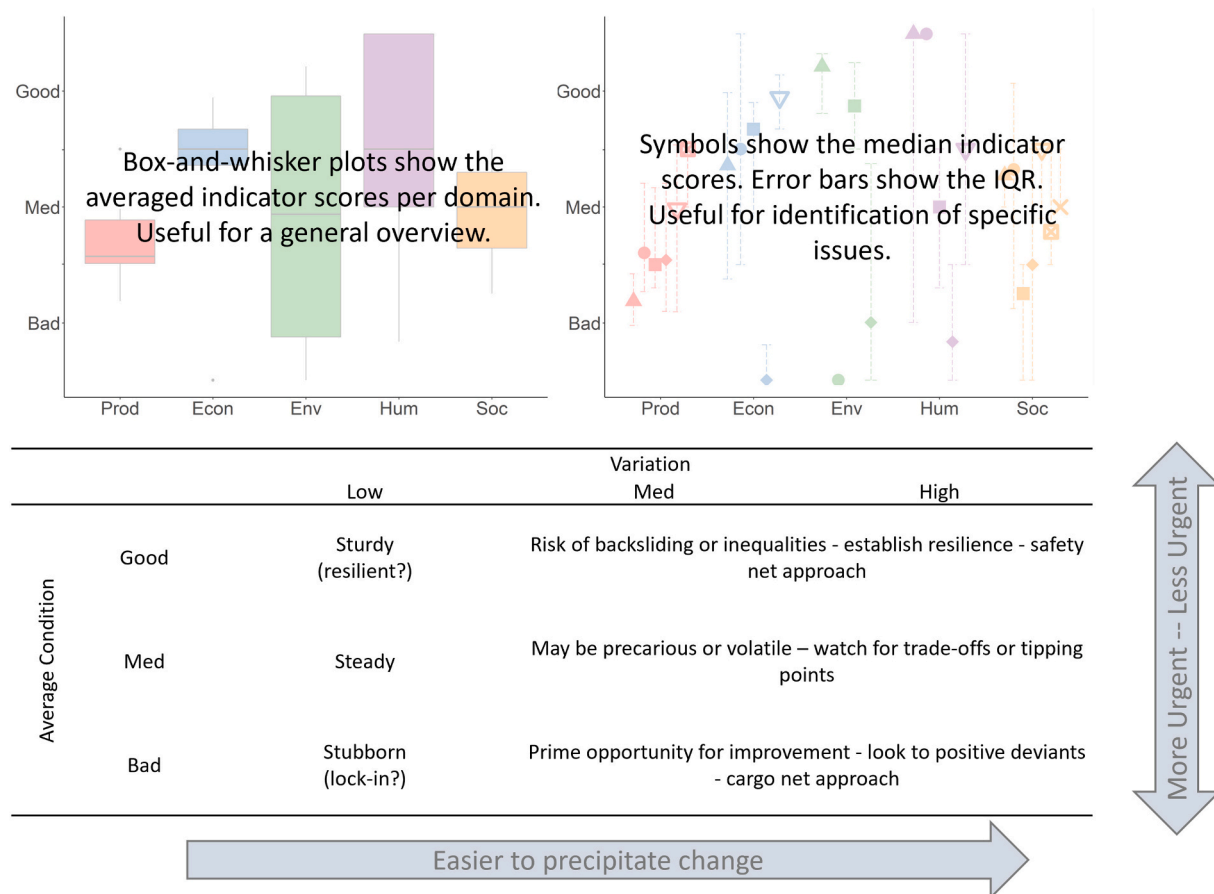


Fig. 2. A guide for how to interpret the sustainable intensification assessments presented below. The boxplot in the upper left of the figure shows the condition of five “sustainability domains”, which are: agricultural production, economic, environment, human welfare, and social. The point plot in the upper right of the figure shows the median and inter-quartile range of each of the indicators monitored. The table in the lower part of the figure provides framework for prioritisation of the issues covered by the indicators. Indicators with worse overall status may be considered more urgent, and indicators with low variation within the site may be considered more challenging to influence. This can be “good” in the case of high-scoring indicators, implying resilience, but “bad” in the case of low-scoring indicators, implying lock-in or intractable issues. A “safety net approach” means protecting the gains made, and would be more suitable for issues where many households are doing well; whereas a “cargo net approach” means lifting up the population on issues where many households fare poorly.

environment (R Core Team, 2021).

3. Results

3.1. Agricultural productivity and sustainability indicators in the study sites

3.1.1. Agricultural productivity

An approximately linear productivity gradient was observed in each site. In Basona Worena the difference between the 10th and 90th percentiles was four-fold, in Endamehonei the difference was five-fold, in Lemo it was four-fold, and in Sinana it was three-fold. The calorie-productivity metric was of the same order of magnitude in each site, although it was highest on average in Sinana and lowest on average in Lemo. The productivity metric for crops was substantially greater than the livestock metric, due to the denominators used (one hectare of land can produce more calories than one large animal). These data and farm-household characteristics are summarised per site and per intensification tercile in Table 3.

As expected, there were large differences between the study sites. Comparisons were made between productivity terciles within each site. There were small but significant differences in land area, and larger significant differences in gross farm income. The more productive farms had slightly smaller land area compared to the less productive farms. There was a general pattern, significant in two of the sites, that more productive farms adopted more of the promoted technologies. The asset base (including off farm income) was similar between the terciles, which implies that the households in the upper terciles produced more with a similar quantity of productive resources, thus indicating intensification.

Table 3

Farm characteristics, by productivity tercile and study site. Productivity terciles were calculated according to the sum of crop kilo-calories (kCal) of production per hectare and livestock kCal production per tropical livestock unit (TLU) (columns 4 and 5). Values presented are means, with standard deviation in parentheses. Significant differences between productivity terciles within a site are indicated by differences in superscript letters, according to Wilcoxon Rank Sum Test, $p < 0.05$.

Site	Productivity tercile	n	Crop productivity (kCal ha ⁻¹ yr ⁻¹)	Livestock productivity (kCal TLU ⁻¹ yr ⁻¹)	Hs'hold members	Crop land (ha)	Lstk owned (TLU)	Gross farm income (USD yr ⁻¹)	Off farm income (USD yr ⁻¹)	New techs in use
Basona Worena	Least	48	2.0E+06 (6.18E+05)	2.0E+05 (2.76E+05)	4.2 ^a (1.6)	1.3 ^a (0.6)	1.8 ^a (1.3)	1040 ^a (4170)	68 ^a (1114)	1.7 ^a (1.2)
	Mod.	48	3.5E+06 (5.80E+05)	5.1E+05 (4.49E+05)	4.3 ^a (1.4)	1.4 ^a (0.5)	1.9 ^a (1.0)	1623 ^b (1868)	484 ^a (1493)	2.8 ^b (1.7)
	Most	47	6.6E+06 (7.88E+06)	3.9E+05 (1.02E+06)	4.3 ^a (1.7)	1.0 ^a (0.6)	1.9 ^a (1.1)	2514 ^b (3886)	629 ^a (3214)	2.5 ^b (1.5)
Endam-ehonei	Least	71	2.9E+06 (1.01E+06)	1.9E+05 (4.21E+05)	5.4 ^a (1.8)	0.8 ^a (0.5)	1.0 ^a (1.1)	835 ^a (827)	126 ^a (493)	2.2 ^a (1.4)
	Mod.	71	5.7E+06 (1.03E+06)	3.3E+05 (6.34E+05)	5.5 ^a (2.1)	0.7 ^a (0.3)	1.3 ^a (1.3)	1432 ^b (1270)	594 ^a (1930)	2.6 ^a (1.4)
	Most	70	9.9E+06 (5.51E+06)	1.3E+06 (5.12E+06)	5.6 ^a (1.8)	0.7 ^b (0.3)	1.3 ^a (1.5)	1971 ^c (2625)	354 ^a (1413)	2.8 ^a (1.8)
Lemo	Least	85	1.8E+06 (5.58E+05)	1.5E+05 (3.19E+05)	6.6 ^a (2.1)	1.5 ^a (0.9)	0.7 ^a (0.5)	279 ^a (370)	26 ^a (409)	1.7 ^a (1.3)
	Mod.	84	3.0E+06 (5.03E+05)	2.6E+05 (4.06E+05)	6.7 ^a (2.0)	1.3 ^b (0.7)	0.9 ^b (0.5)	351 ^a (1911)	9 ^a (83)	1.5 ^a (1.3)
	Most	84	4.9E+06 (1.62E+06)	1.1E+06 (2.32E+06)	6.7 ^a (1.9)	1.1 ^c (0.7)	0.7 ^a (0.4)	606 ^b (805)	28 ^a (505)	1.6 ^a (1.2)
Sinana	Least	54	4.9E+06 (1.48E+06)	3.7E+05 (4.79E+05)	6.7 ^a (2.4)	3.5 ^a (2.0)	1.3 ^a (0.9)	3124 ^a (3543)	133 ^a (433)	1.1 ^a (0.7)
	Mod.	54	8.2E+06 (1.49E+06)	6.2E+05 (9.03E+05)	7.1 ^a (2.2)	3.5 ^a (2.5)	1.4 ^a (1.0)	5017 ^b (5361)	281 ^a (2249)	1.7 ^b (1.1)
	Most	53	1.3E+07 (4.27E+06)	7.8E+05 (1.57E+06)	6.9 ^a (2.3)	3.6 ^a (2.3)	1.5 ^a (1.3)	8826 ^c (7260)	266 ^a (5215)	1.4 ^b (0.8)

3.1.2. Productivity and the aggregated sustainability domains

Correlation tests between productivity metrics and the sustainability domain scores revealed significant positive correlations (Table 4). For all sites combined, there were significant positive correlations between productivity and each of the sustainability domains. These were strongest for the economic domain, and weakest for the environmental and production domains. Per study site, the findings were similar, although no significant correlations were found for the environmental domain, while the production domain showed stronger correlations per site than overall (perhaps influenced by the very different agro-ecologies in each site). In Basona Worena the production, economic, human, and social domains were all significantly correlated with productivity (with the strength of correlation in the order listed). In Endamehonei, the production, economic, and human domains were significantly correlated with productivity. In Lemo the correlations were weaker, but still significant for the production, economic, and human domains. In Sinana the correlations were significant for the production, human, economic, and social domains. The domain scores per site and intensification

Table 4

Correlation between farm productivity and aggregated sustainability domain scores. Statistically significant correlations are indicated by bold text ($p < 0.05$). Note that productivity metrics were excluded from the production domain for these tests.

	Production	Economic	Environmental	Human	Social
All sites	0.1	0.4	0.1	0.2	0.2
Basona Worena	0.4	0.3	0.1	0.2	0.2
Endamehonei	0.4	0.4	0.0	0.3	0.1
Lemo	0.2	0.2	0.1	0.2	0.0
Sinana	0.3	0.2	0.0	0.3	0.1

tercile are visualised in Fig. 3, which shows that generally average domain scores did not change much along the productivity gradient, but that there was often a noticeable reduction in the spread of domain scores.

3.1.3. Productivity and the sustainability indicators

Correlations between productivity and the re-scaled indicator scores are presented in Table 5. In the production domain, staple crop yield was

most strongly correlated with productivity, which is logical as crop productivity was the main constituent of the intensification metric, and the most popular interventions aimed to improved grain yields. The correlation between milk yield and productivity was weaker, significant only in Basona Worena, where livestock were the most important for livelihoods. More unexpectedly, the crop and livestock diversity metrics showed weak positive correlations, suggesting a degree of production intensification through diversification. In the economic domain, the two

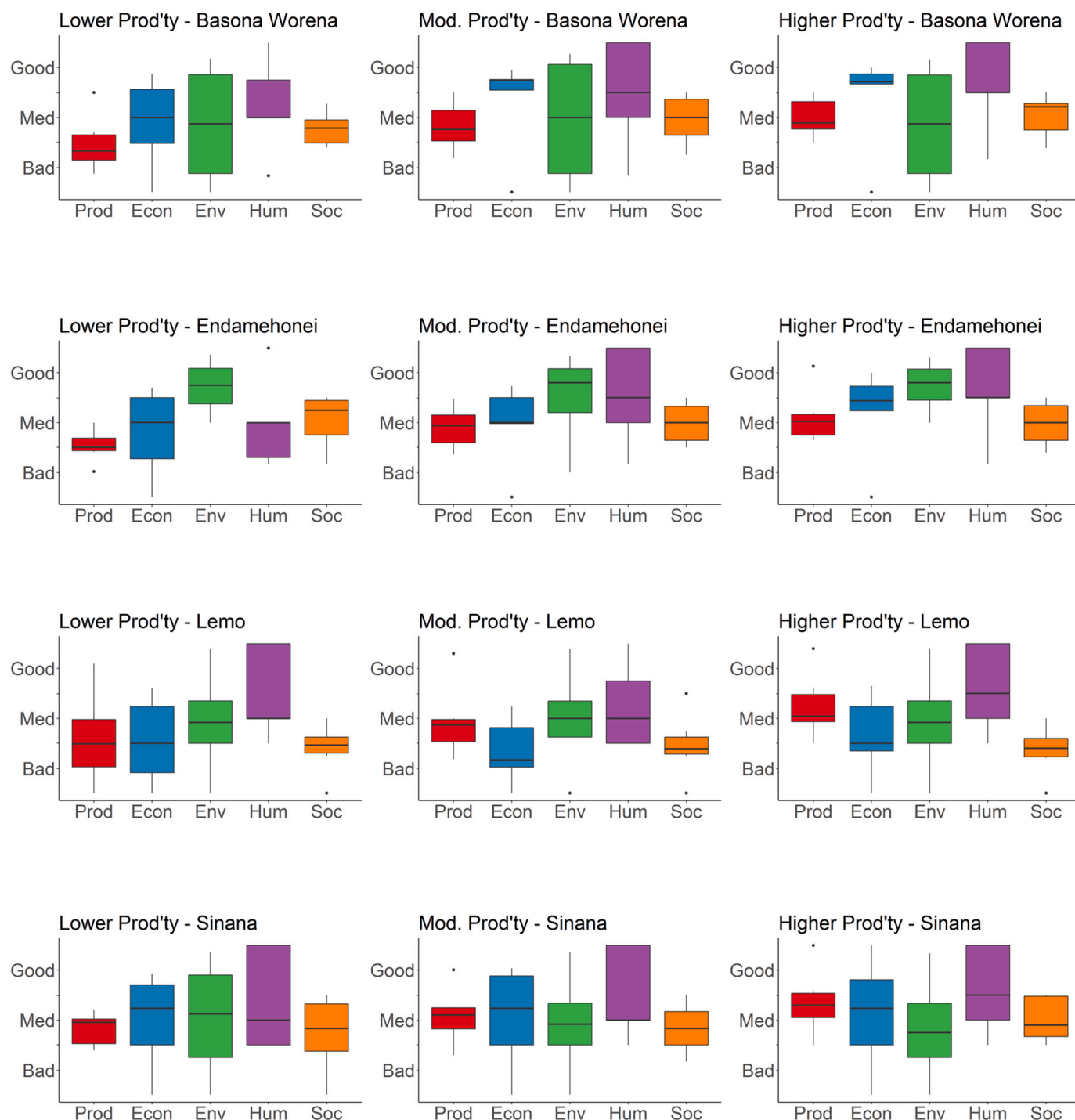


Fig. 3. The aggregated scores for each of the five sustainability domains: agricultural production, economic, environmental, human welfare, and social. Each row represents one of the four study sites. Within each site, the surveyed population was split into tertiles of agricultural productivity (combining crop and livestock productivity); each column shows one productivity tertile. Productivity metrics were excluded from the production domain for these figures. Looking across a row from left to right, one may see how the sustainability domain scores change as agricultural productivity increases.

Table 5

Correlation between household indicator scores and agricultural productivity. Statistically significant correlations are indicated by bold text ($p < 0.05$). The indicators for crop productivity and livestock productivity were excluded as they were used to calculate productivity metric.

Domain	Indicator	Basona Worena	Endamehonei	Lemo	Sinana
Production	Staple crop yield	0.7	0.6	0.5	0.7
	Crop productivity	–	–	–	–
	Crop diversity	0.1	0.1	0.1	–0.1
	Milk yield	0.3	0.1	0.0	0.1
	Lstk productivity	–	–	–	–
Economic	Lstk diversity	0.0	0.1	0.1	0.0
	Total value activities	0.3	0.5	0.4	0.4
	Income sources	0.2	0.3	0.1	0.0
	Market orientation	0.1	0.1	0.1	–0.1
	Off farm income	0.1	0.1	0.1	0.0
Environment	PPI (asset score)	0.0	0.1	0.0	0.1
	GHG emissions	0.0	–0.1	0.0	–0.1
	Irrigation	0.2	0.1	–0.1	0.0
	Land conservation	0.1	–0.1	0.1	0.0
	Soil quality	0.0	0.0	0.1	0.0
Human	HFIAS (hunger)	0.1	0.2	0.1	0.2
	Months food secure	0.1	0.2	0.1	0.1
	Diet diversity	0.2	0.3	0.2	0.2
	Education	0.0	0.1	0.1	0.2
	Novel practices trialled	0.2	0.1	0.0	0.1
Social	Female owned assets	0.0	0.0	–0.1	0.0
	Female income control	0.1	0.1	0.1	0.0
	Dependency ratio	0.0	–0.1	–0.1	0.0
	Group membership	0.1	–0.2	0.0	0.1
	Extension services	0.1	0.1	0.0	0.2
	Skill sharing	0.3	0.3	0.1	0.2
	Gifts and exchange	0.0	0.0	0.0	0.0

variables most significantly correlated to productivity were total value of activities (which is the product of total farm production by the market price of each commodity, plus any off-farm income), and the number of income sources. This suggests that households who intensified not only produced more but also diversified their incomes. The environmental domain showed the weakest and fewest significant correlations. Irrigation use increased significantly in Basona Worena with productivity, otherwise there were non-significant correlations with reduced greenhouse gas emissions in two sites, and increased use of land conservation practices in two sites. Soil quality and productivity were not correlated. In the human domain, there was good evidence for correlation between productivity and food security, as all three indicators returned positive correlations in all sites, and were significant in the majority of cases. The indicators for education and innovation also returned positive correlations with productivity, although these were not significant in most cases. In the social domain the indicators around knowledge exchange returned significant correlations to productivity, in particular peer-to-peer knowledge exchange. The indicators around female control of assets and incomes did not show significant correlations, but were weakly positive in most sites for female control of incomes. The analysis of individual indicators showed that in most cases differences in the domain scores were due to differences in multiple indicators, rather than being

driven by a large difference in a single indicator.

3.2. Site diagnostics

In this section the overall condition of broadly-defined sustainable intensification in each of the four study sites is summarised. The median scores for the 27 indicators are presented in Table 6, using the original units and scales for each indicator. The re-scaled indicators and aggregated sustainability domains are then presented in visually per site in Figs. 4–7.

3.2.1. Basona Worena (Amhara)

The domain score summary in the top left of Fig. 4 shows that the agricultural productivity domain scores the lowest, followed by the environmental and the social. The economic domain looks to be the strongest, and the human domain also to be in generally good condition. Examination of the indicators scores (top right of Fig. 4) presents a slightly more complex situation, where the human domain in particular showed some low scores. Education was poor, and there was a high degree of variation in two of the food security indicators (HFIAS, dietary diversity). However, the observation that agricultural productivity warrants particular attention is supported; crop-related indicators were poor, although livestock related indicators were somewhat better. In the economic domain, off-farm income opportunities were lacking but otherwise indicators were in a good condition. The environmental domain indicators are at the two extremes, with greenhouse gas emissions and land conservation practices in a very good state, and soil quality and irrigation in a very poor state. It may be that over time the land conservation measures will lead to improved soil and water retention. Extension services and skill sharing were moderately good, though group membership was poor on average and highly variable. Looking at the prioritisation table in Fig. 4, the indicators in worst condition related to agricultural production, either directly (e.g. staple crop yield) or indirectly (irrigation, off-farm income to invest in intensification). Many of the human and social indicators were categorised as having “high variation”. Although the average condition (score) may have been acceptable, there were many households in the study site doing much worse than the average. This suggests the need for continued effort to disseminate the activities and benefits within the site.

3.2.2. Endamehonei (Tigray)

In Endamehonei agricultural production was the poorest scoring domain (Fig. 5). Staple crop yields and crop diversity were low. Total crop productivity was in a better condition than Basona Worena or Lemo, which may suggest better soil conditions. Livestock productivity was highly variable. The economic domain was in a moderate condition, with few opportunities for off-farm income and high variation in the total value of activities. The environmental domain appeared to in relatively good condition, with greenhouse gas emissions and land conservation practices in a good state with low variation, and average irrigation and soil quality in a moderate condition, although with very high variation. The human domain again showed a high degree of variation, and although the food security indicators were good on average, HFIAS and dietary diversity showed high variation. Education levels were generally poor. In the social domain knowledge exchange was moderately good. According to the prioritisation table, the most urgent issues are staple crop yields, crop diversity, off farm income, and education. High variation in livestock productivity, food security, knowledge exchange, total value of production, and soil quality suggests that higher scores on indicators are possible within the site, and that distribution or access issues are a greater barrier than technical feasibility.

3.2.3. Lemo (SNNPR)

The summary of the Lemo site (Fig. 6) showed sustainable

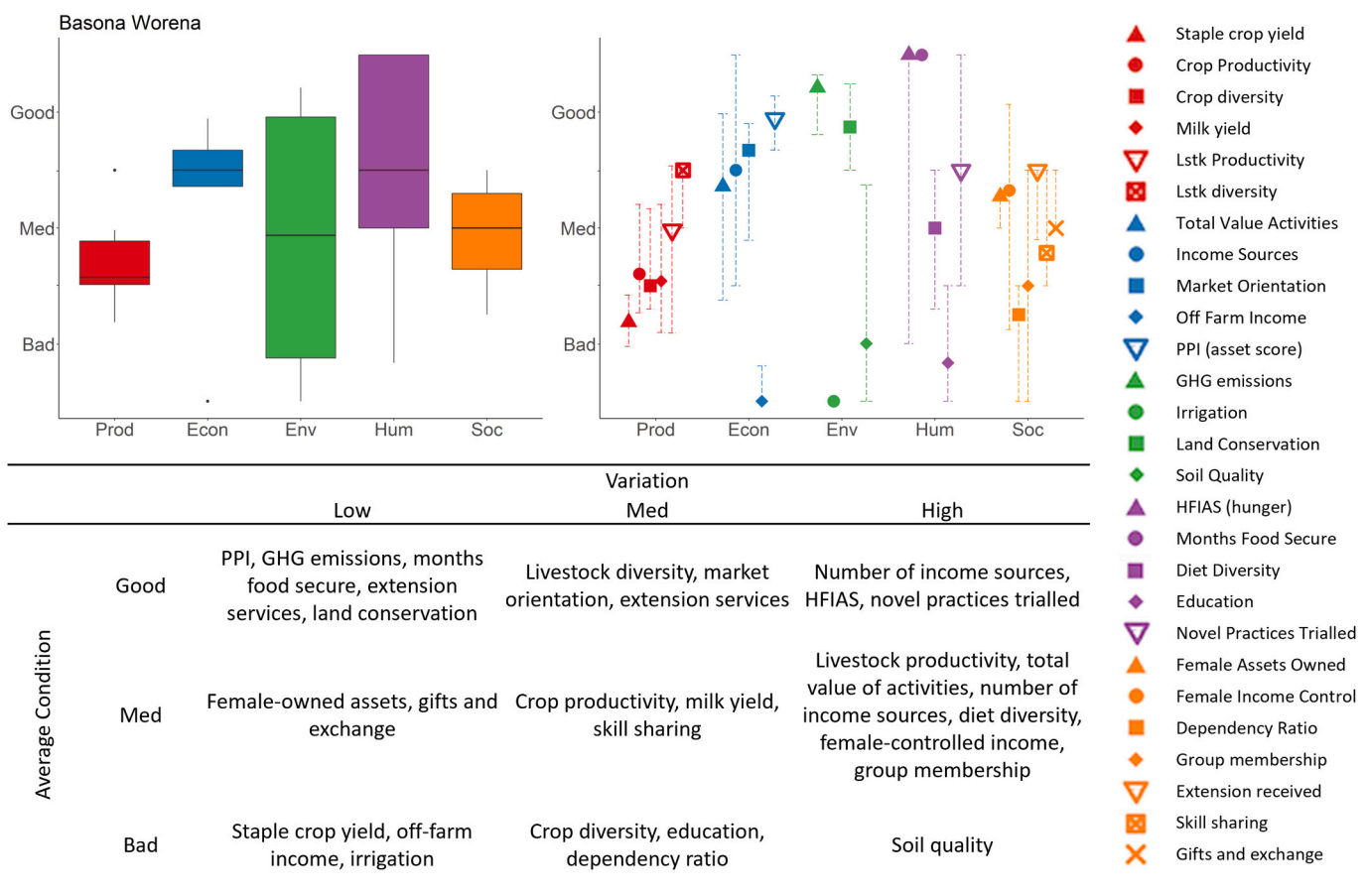
Table 6

Median indicator scores expressed in their original units, for each of the study sites.

Domain	Indicator	Basona Worena	Endamehonei	Lemo	Sinana	Unit
Production domain	Staple crop yield	1.6	2.0	1.6	3.1	t ha ⁻¹ yr ⁻¹
	Crop productivity	3.41E+06	5.59E+06	2.78E+06	8.36E+06	kCal ha ⁻¹ yr ⁻¹
	Crop diversity	5	5	12	4	Species grown
	Milk yield	2.4	1.5	1.5	1.5	Litres cow ⁻¹ day ⁻¹
	Lstk productivity	2.82E+05	2.27E+05	2.33E+05	3.92E+05	kCal TLU ⁻¹ yr ⁻¹
Economic domain	Lstk diversity	4	3	3	3	Species kept
	Total value of activities	1.8	1.4	0.6	3.5	USD person ⁻¹ day ⁻¹
	Income sources	5	5	3	3	Count
	Market orientation	65	71	23	62	% produce sold
	Off farm income	0	0	0	0	USD household ⁻¹ yr ⁻¹
Environment domain	PPI (asset-based poverty estimate)	86	68	68	68	% chance out of poverty
	GHG emissions	0.6	0.3	0.2	0.3	tCO _{2eq} pers. ⁻¹ yr ⁻¹
	Irrigation	0	3	0	0	Months yr ⁻¹
	Land conservation	6	6	2	2	Count of practices
	Soil quality	1	2	2	2	Farmer perception (0–3)
Human domain	HFIAS (hunger)	4	4	4	4	Severe to no experience of hunger (1–4)
	Months food secure	12	12	12	12	Months yr ⁻¹
	Household dietary diversity	5	5	5	5	Food groups consumed weekly (0–10)
	Education	2	3	4	4	Ordinal (1–7)
	Novel practices trialled	3	3	2	1	Count over last 5 years
Social domain	Female assets owned	2.8	2.8	2.2	2.3	Assets owned or co-owned by females (0–4)
	Female control	32	36	11	33	% of livelihood controlled by females
	Dependency ratio	55	95	68	81	Ratio of non-workers:workers
	Group membership	1	2	1	2	Group membership yr ⁻¹
	Extension services	9	9	3	5	Frequency and quality (1–18)
	Skill sharing	5	4	0	2	Count of peers receiving skills
	Gifts and exchange	0	1	1	1	Count of substantial items given or received yr ⁻¹

intensification to be in a moderate-to-poor condition. Particularly low-scoring was the social domain, with knowledge exchange indicators and gender equity indicators in poor condition. The economic domain

was also in poor condition, with the total value of activities, number of income sources, and off-farm income all scoring poorly and with little variation. Staple crop yields and total crop productivity were poor, and

**Fig. 4.** Sustainable intensification diagnostics for the Basona Worena (Amhara) study site.

livestock production indicators in moderate condition, suggesting that crop interventions should be prioritised. In the environmental domain, use of land conservation practices was notably lower than in Basana Worena or Endamehonei, and there was a high degree of variation in soil quality. In the human domain, food security indicators were again good on average but showed high variation. The priority issues for further intervention appeared to be increasing income generation, crop yields, and knowledge exchange, and with a focus on female empowerment.

3.2.4. Sinana (Oromia)

The diagnostics of the Sinana site showed that all domains scored in moderate condition, with high variation in the economic, human, environmental domains, and a low degree of variation in the production and social domains. Crop productivity was good, and staple crop yields were moderate and with low variation (which is better than observed in other sites). Crop diversity was the only production indicator which scored poorly. The economic indicators of total value of activities and market orientation were in a very good state, and the asset-based poverty score was good. The number of income sources was on the low side, which makes sense considering the predominant grain monocropping in Sinana. As in other sites, there were few opportunities for off-farm income, soil quality was highly variable, and per capita greenhouse gas emissions were very low. Food security was good on average but showed high variation. In terms of knowledge exchange, group membership was moderately good, but extension services and peer-to-peer exchange were poor. The share of female-controlled income was relatively high in Sinana compared to other locations. The priorities for this site would be to focus on agricultural and economic diversification, knowledge exchange, and to further improve soil fertility.

4. Discussion

In contrast to the majority of studies on the implementation of sustainable intensification (Weltin et al., 2018; Thomson et al., 2019) this study did not focus at the field scale, on individual technologies/practices, and did not focus primarily on environmental outcomes. We assessed the outcomes from a wide range of sustainable intensification technologies which had been concurrently implemented in an uncontrolled manner, focussing our analysis at the farm-to-community scale, and considering the multi-functional outcomes of agriculture (Binder et al., 2010). For this we integrated a rapid, harmonized household survey tool (Hammond et al., 2017) with a framework approach to assess broadly-defined sustainable intensification (Musumba et al., 2017). We applied an expert-led threshold setting and re-scaling approach, to facilitate context specific indicator interpretation, and developed a novel graphical presentation of the results. In this way we addressed the challenge formulated by Weltin et al. (2018) that more studies should focus on higher spatial scales (above plot level), assessment of the simultaneous use of multiple technologies, and improve the context-sensitivity of approaches.

4.1. Trade-offs and synergies, identification and assessment

Agricultural productivity returned positive correlations (synergies) with the agricultural, economic, social, human, and sometimes environmental, domains across all four sites investigated in Ethiopia (Tables 4, 5, and Fig. 2). We did not find evidence of negative correlations (trade-offs) between productivity and any of the sustainability domains. Trade-offs between agricultural productivity and other domains, especially the environment (e.g. Vang Rasmussen et al., 2018), are so

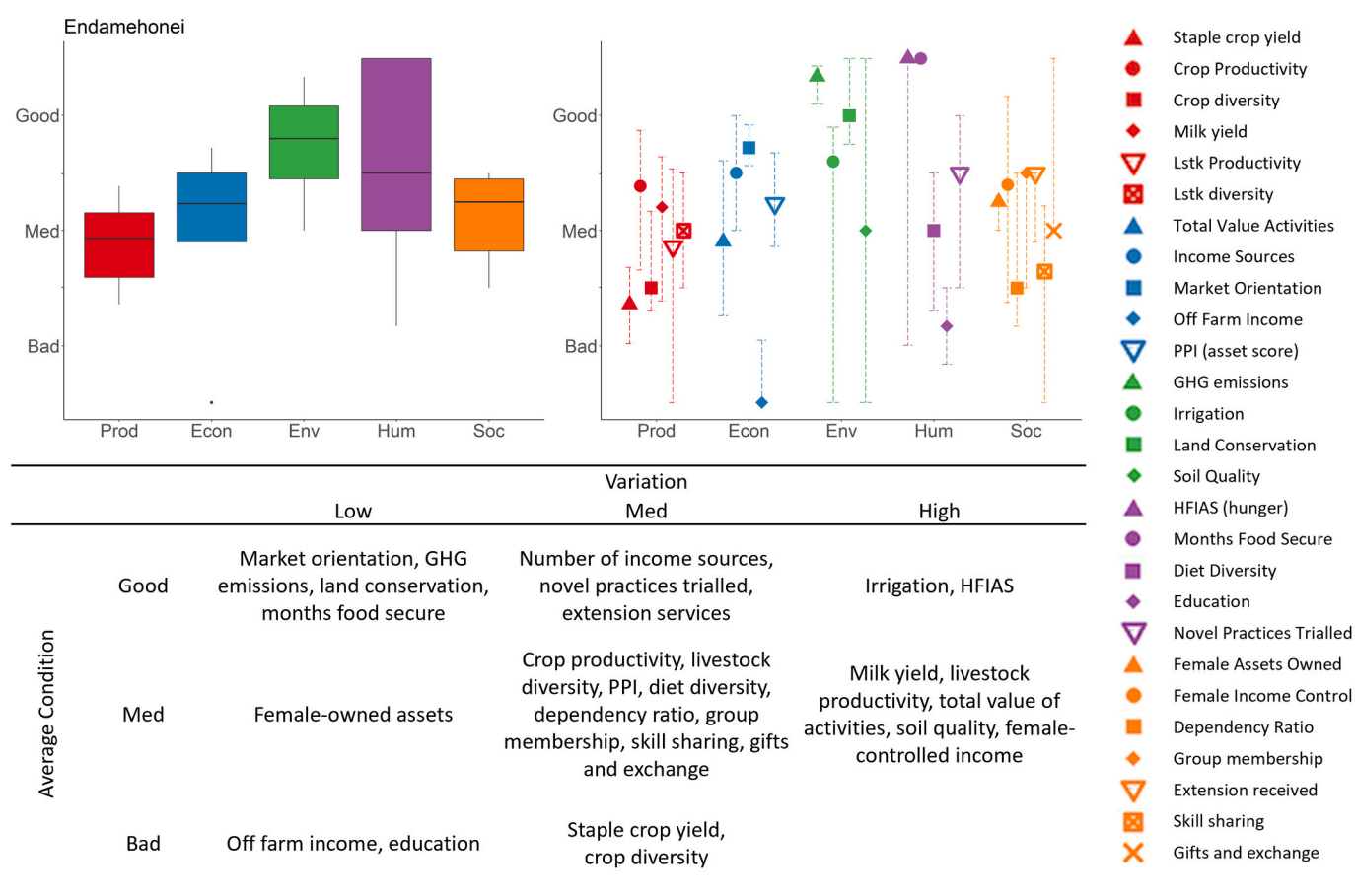


Fig. 5. Sustainable intensification diagnostics for the Endamehonei (Tigray) study site.

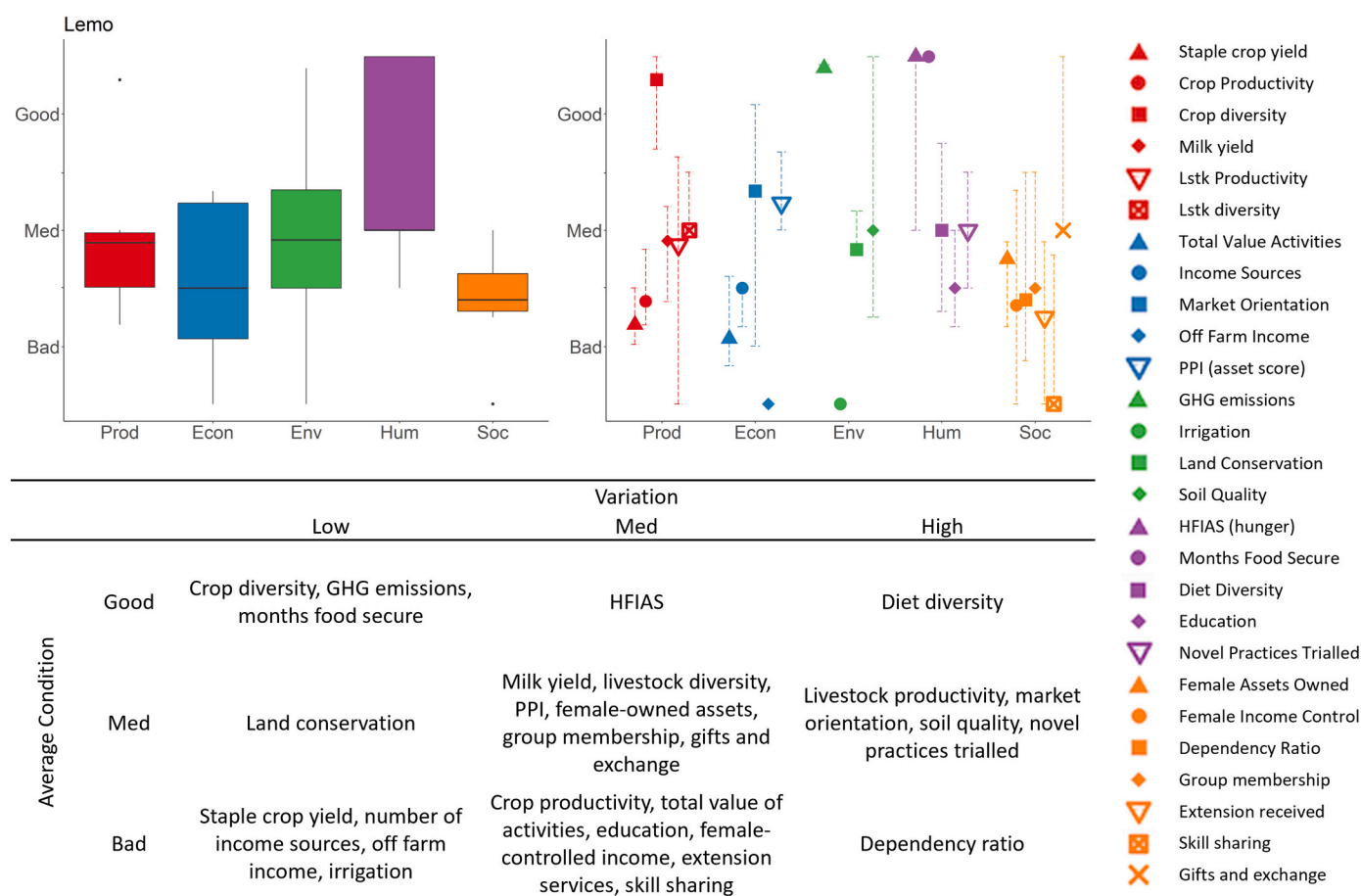


Fig. 6. Sustainable intensification diagnostics for the Lemo (SNNPR) study site.

commonly reported that they may be misunderstood as an inherent externality of intensification. Whilst we observed the least synergy between the environmental domain and agricultural productivity, we did not identify trade-offs. In the economic, human, and social domains, clear synergies arose between agricultural productivity and the indicators underpinning the domains. These results suggest that at such low productivity levels, common in many smallholder systems across sub-Saharan Africa (e.g. [Tittonell and Giller, 2013](#)), higher productivity goes hand-in-hand with better performance across multiple dimensions of sustainability. Key trade-offs are only likely to occur at productivity levels which require higher levels of inputs (such as agro-chemicals, high levels of irrigation) which would affect the environmental indicators as well as possibly equity-based social and human welfare indicators.

Another reason for the absence of trade-offs in our assessment may have been the selection of locally appropriate and complementary technologies. The Africa RISING program in Ethiopia was unusual in the degree of effort put into situation analysis, participatory technology selection, and the enhancement of knowledge exchange capacities between stakeholders ([Pound et al., 2015](#)). Such approaches have been recognised as important in moving beyond technical demonstrations and achieving widespread uptake, both for farmers ([Jambo et al., 2019](#); [Marinus et al., 2021](#)) and for extension agents ([Jiao et al., 2019](#); [Ortiz-Crespo et al., 2020](#)). This is also supported by the observation that indicators for knowledge exchange were generally correlated with increased productivity. The number of technologies available, combined with the freedom for farmers to select as they wished, may have allowed the participant farmers to balance trade-offs themselves. A final possible cause is that the assessment methodology was not sufficiently detailed at the necessary level of granularity to identify the relevant trade-offs. This

topic is discussed in greater detail below, and we conclude that it does not undermine the findings.

Confidence in our findings was increased by comparison to another assessment of agricultural sustainability in the same four regions ([Mutyasira et al., 2018](#)). Mutyasira et al. used different indicators to measure sustainability and a different analysis methodology to identify correlates with sustainability outcomes, but their findings were coherent with ours. They found the same ranking of the sites in terms of overall sustainability scores as reported in this manuscript; and found that most farms were in a poor to moderate condition, although a significant minority of farms scored well.

We chose to focus on the identification of trade-offs or synergies between agricultural production and the other sustainability domains. This was inspired by the question implicit in the juxtaposition of the terms “sustainable” and “intensification”. However, many other potential trade-offs exist within the farm system (e.g. whether to specialise or diversify income streams), within communities (e.g. whether to protect watersheds) and beyond (e.g. interaction with markets). Promising methodologies to explore trade-offs and the role of decision making include integrated assessment models, market equilibrium models ([Valdivia et al., 2012](#) integrated both), and agent based models ([Rouleau and Zupko, 2019](#)). Assessment of the sustainability of different farm typologies within a single landscape is an interesting angle ([Haileslassie et al., 2016](#)), and raises the question of the level at which analysis should be directed (farm, farm type, community, landscape).

Location and scale are also relevant in the identification of trade-offs and synergies (see e.g. [Goldstein et al., 2012](#)). Spatial assessment requires a spatially explicit sampling strategy as well as spatially explicit evaluation of indicator scores. The impact of an action can vary according to the location at which it is performed within a landscape: for

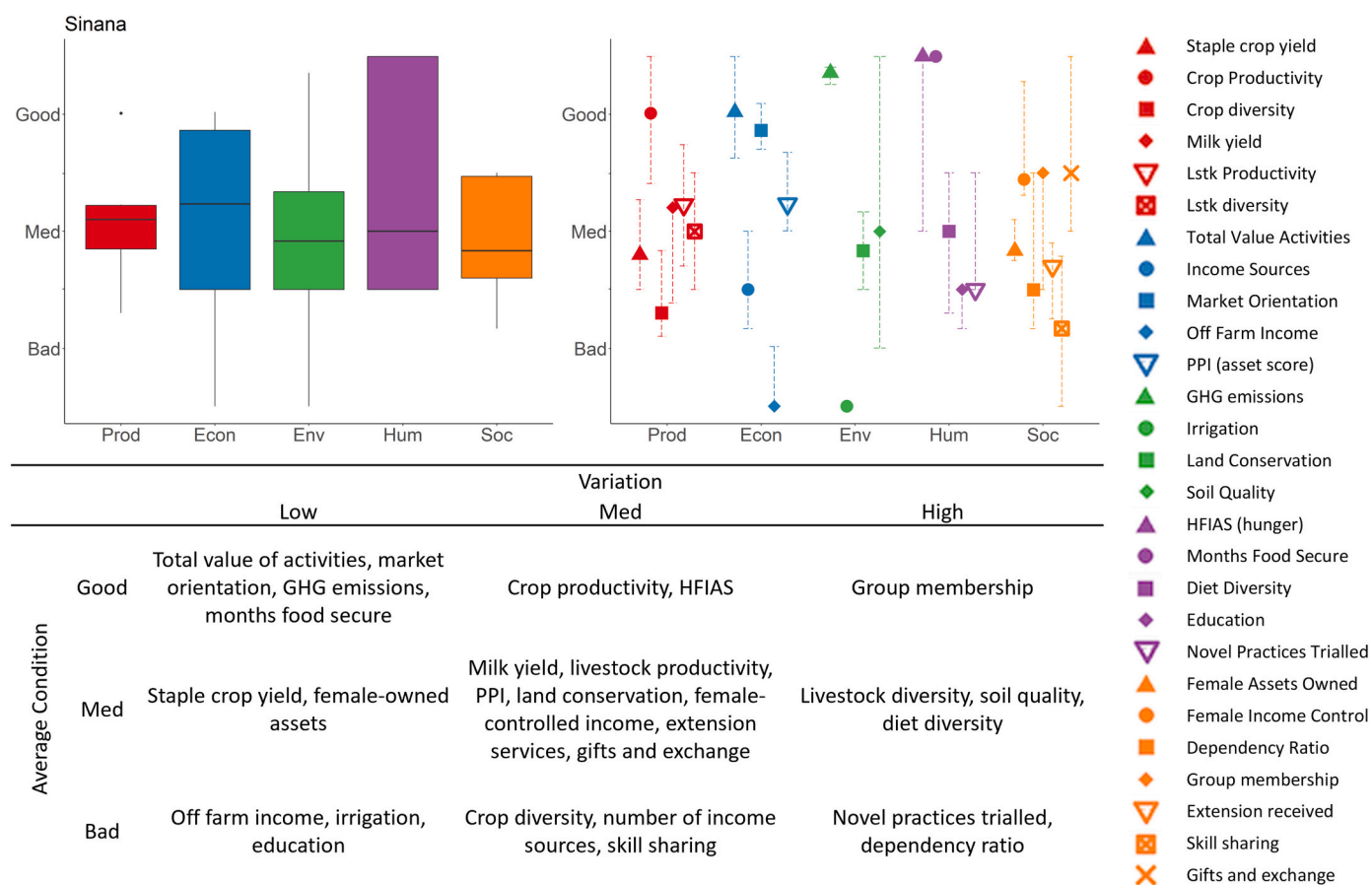


Fig. 7. Sustainable intensification diagnostics for the Sinana (Oromia) study site.

example deforestation in a flat valley bottom typically causes less disruption to the hydrology compared to deforestation of steep slopes in the same watershed. Similarly the scale at which the action is carried out can change the outcome: deforestation on a landscape scale can have severe impacts on biodiversity and weather patterns, but deforestation on single farm scale usually does not. The integration of these issues into agricultural sustainability assessments remains uncommon (Inwood et al., 2018).

4.2. Inherent bias in sustainability assessments

Many authors observe that indicator selection and index creation are particularly subjective, value-laden, and inherently political processes (Asokan et al., 2020; Acosta-Alba and Van der Werf, 2011; Gan et al., 2017). Selection or omission of indicators, weighting of indicators, and aggregation of indicators into indices can obfuscate or emphasise certain findings, depending on the decisions made. In order to engender trust in the findings transparency of process is paramount, and the presentation of both aggregated and disaggregated indicator values is recommended (Bezlepikina et al., 2011; Inwood et al., 2018). The indicators chosen should be sufficient for stakeholder decision making and capture an adequate degree of system complexity, whilst they should also be sufficiently simple to be routinely monitored, readily understood, as well as responsive to stresses within the system and sensitive enough to detect differences between households (Dale and Beyeler, 2001; Inwood et al., 2018). The degree of indicator score variation observed in this study suggests that differences between households were generally detected, and therefore the indicators sufficiently sensitive. Kanter et al. (2016) argue for selecting a broad base of indicators and note that there are few holistic datasets that permit an analysis linking sustainability domains in

agriculture. In our study, the environmental indicators were the weakest of the five domains in this regard, although the soil and water indicators are of key importance to the Ethiopian systems of interest. There were notable omissions for biodiversity, environmental health, and pollutants. This could be improved by conducting a more profound participatory assessment of each issue (for soil quality see Barrios et al., 2012) or through biophysical monitoring. However the lack of well-validated score-card type approaches for the rapid assessment of environmental issues (e.g. Barrios and Mortimer, 2014) is an impediment, which has been overcome for other topics such as food security (e.g. Coates et al., 2007).

Index creation is essentially an interpretive process whereby numerous indicators are compiled for comprehension and communication purposes (Singh et al., 2012). Dimension reduction is the outcome, but the purpose is to increase *meaningfulness* of the findings (Pollesch and Dale, 2015; Gan et al., 2017). Depending on the number of indicators, the number of dimensions they will be aggregated into, and the degree of compensation permitted, different mathematical (Pollesch and Dale, 2015; Gan et al., 2017), interpretive (Gerrard et al., 2011; Kanter et al., 2016), or visual (Miettinen, 2014) methods might be appropriate. The presentation of disaggregated indicator values alongside the aggregated domain scores avoided many of the problems inherent in sustainability assessments which only present aggregated values. The decision to apply no weighting to the indicators was taken partly due to lack of credible weights, and partly as unweighted indicators are more intuitively comprehensible, thus inviting readers to weight importance for themselves using the disaggregated indicators (Figs. 4 to 7). Clarity over the reference value for each indicator – the value against which an indicator score is judged – is also necessary and often glossed over (Acosta-Alba and Van der Werf, 2011). We made the reference values

explicit (Table 2) and used them for re-scaling indicators to aid in interpretation. We signposted the subjectivity by use of the terms “good”, “medium”, and “bad” to describe indicator scores. This is an extension of the widely used traffic-light approach (e.g. Gerrard et al., 2011; Anderson et al., 2015), and a similar approach has recently been used for the evaluation of agro-ecological intensification (FAO, 2019). We used thresholds supported by peer-reviewed literature wherever possible, but in some cases there was no option but to use relative thresholds. An interesting outcome of the aggregation function applied is that there were stronger correlations to productivity at domain level compared to indicator level. This could be due to the sum of many small indicator improvements which alone were smaller or not significant but combined led to significant domain level improvements.

The selection of the metric for agricultural productivity was another issue of key sensitivity. Calorie output per unit of land and per unit of livestock is meaningful in terms of overall food output efficiency, but excludes inputs such as labour, chemicals, and natural resources. To a limited degree, the other indicators cover these issues. Lifecycle assessment of these issues would be a more comprehensive approach (Garnett, 2014), although challenging in data sparse environments. We also trialled cash value per land and livestock unit as a productivity metric. This increased the contribution of livestock to farm productivity (as livestock products are typically higher value compared to crops), but the overall findings of the study were similar. One notable difference was that the female asset ownership and female income control were more strongly and significantly linked to intensification when the intensification metric used was cash value. This may be a result of the Africa RISING program efforts to involve females in cash generating activities (Mulema et al., 2019). However, on balance we rejected using cash value as an intensification metric as it blurred the distinction between the economic and production domains.

Communication of sustainability assessments often relies heavily on visualisation and summary tables. Striking the appropriate balance between succinct and intuitive conveyance of relevant information, acknowledging uncertainty, and avoiding misdirection and extraneous detail is not trivial (Bezlepikina et al., 2011; Miettinen, 2014); but is crucial to support well-informed decision making (Kanter et al., 2016). Stakeholder engagement is an essential step towards impact (Klapwijk et al., 2014), in order to contextualise the results (Weltin et al., 2018) and identify appropriate management responses (Inwood et al., 2018). The degree to which stakeholders can make sense of the results and the degree to which they choose to act upon them depend in part on the decisions made in earlier stages of the study design process (de Olde et al., 2016).

4.3. Avenues for improvement

Although this manuscript addressed some gaps identified in reviews of agricultural sustainability assessments many of the aspirations described by authors of such reviews remain unfulfilled. For example, recent reviews have found that sustainability was generally framed in a narrow sense with too few indicators used (Reich et al., 2021; Chopin et al., 2021); that assessments tended to be top down and lack participation (Klapwijk et al., 2014; Chopin et al., 2021); that there was insufficient effort to distinguish impacts from drivers (Chopin et al., 2021); and that time, space, and scale interactions have not been adequately incorporated into research designs (Inwood et al., 2018). The approach used here could be bolstered to address some of these remaining gaps. In particular, a research design making use of spatially-weighted sampling for higher-frequency surveys, linked to environmental point measurements and remote sensing could deliver a sustainability assessment at the landscape level. The assessment of trade-offs and synergies could be developed from identification to spatially- and temporally-explicit prediction, thus leading to scenario analyses and policy-relevant advice.

5. Conclusions

We presented a novel application of the Sustainable Intensification Assessment Framework in an integrated manner consisting of data collection, indicator quantification, interpretation, and communication. We assessed the combined outcome of a wide variety of sustainable intensification technologies adopted in an uncontrolled manner by farmers. The most important finding was an absence of trade-offs between agricultural productivity and the net economic, human welfare, social, and environmental outcomes. On the contrary, in the majority of cases significant synergies were found between increased agricultural productivity and especially the economic, human and social domains. We conclude that the technologies promoted (and the enabling environment established) were suitable to enhance the low levels of agricultural productivity whilst also enhancing other sustainability outcomes.

However, in each of the four study sites agricultural productivity remained below attainable potential. Further support measures to increase both crop and livestock productivity are still required. In addition to technical agricultural support, development of the enabling environment should also be supported. Education of household heads was generally very low, and there was very little generation of off-farm income. Access to good quality extension and knowledge exchange opportunities (e.g. peer-to-peer learning) was associated with increased agricultural production. In Basona Worena and Endamehoni further efforts should be put into disseminating the good practices which were evident. In Lemo further work on building social capital and the enabling environment may unlock productivity and income gains. In Sinana crop production and incomes were relatively high, but there was relatively little innovation, implying that the farmers did not view the promoted technologies as highly suitable for their grain-oriented farming system.

In summary, this manuscript provided evidence that progress towards a broadly-conceived notion of sustainability was possible through the intensification of agriculture on small farms in the Ethiopian highlands, and that the farmers demonstrated an appetite to take part in this process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was undertaken with support from Africa RISING, a program financed by the United States Agency for International Development (USAID) as part of the United States Government's Feed the Future Initiative. We also acknowledge the support of the CGIAR Research Program on Livestock and its donors. The content is solely the responsibility of the authors and does not necessarily represent the official views of USAID or the U.S. Government or that of the Africa RISING program. Africa RISING is aligned with research programs of the CGIAR.

References

- Acosta-Alba, I., Van der Werf, H.M.G., 2011. The use of reference values in Indicator-based methods for the environmental assessment of agricultural systems. *Sustainability* 3 (2), 424–442. <https://doi.org/10.3390/su3020424>.
- Africa RISING, 2019. Africa RISING Program Overview. <http://africa-rising-wiki.net/File:Ar-programoverview.pdf#file> accessed 03/03/2021.
- Anderson, J.L., Anderson, C.M., Chu, J., Meredith, J., Asche, F., Sylvia, G., et al., 2015. The fishery performance indicators: a management tool for triple bottom line outcomes. *PLoS One* 10 (5), e0122809. <https://doi.org/10.1371/journal.pone.0122809>.

- Asokan, V.A., Yarime, M., Onuki, M., 2020. A review of data-intensive approaches for sustainability: methodology, epistemology, normativity, and ontology. *Sustain. Sci.* 15, 955–974. <https://doi.org/10.1007/s11625-019-00759-9>.
- Barrios, E., Mortimer, P., 2014. NaRIA Scorecard - Natural Resource Integrity Assessment. Available at: <https://stats4sd.org/resources/238>.
- Barrios, E., Coutinho, H.L.C., Medeiros, C.A.B., 2012. InPaC-S: Participatory Knowledge Integration on Indicators of Soil Quality – Methodological Guide. World Agroforestry Centre (ICRAF), Embrapa, CIAT, Nairobi, p. 178 (ISBN: 978-92-9059-329-4).
- Bezlepkin, I., Reidsma, P., Sieber, S., Helming, K., 2011. Integrated assessment of sustainability of agricultural systems and land use: methods, tools and applications. *Agric. Syst.* 104 (2), 105–109. <https://doi.org/10.1016/j.agsy.2010.11.002>.
- Binder, C.R., Feola, G., Steinberger, J.K., 2010. Considering the normative, systemic and procedural dimensions in indicator-based sustainability assessments in agriculture. *Environ. Impact Assess. Rev.* 30, 71–81.
- Cassman, K.G., Grassini, P., 2020. A global perspective on sustainable intensification research. *Nat. Sustain.* 3, 262–268. <https://doi.org/10.1038/s41893-020-0507-8>.
- Chopin, P., Mubaya, C.P., Descheemaeker, K., et al., 2021. Avenues for improving farming sustainability assessment with upgraded tools, sustainability framing and indicators. A review. *Agron. Sustain. Dev.* 41, 19. <https://doi.org/10.1007/s13593-021-00674-3>.
- Coates, J., Swindale, A., Bilinsky, P., 2007. Household Food Insecurity Access Scale (HFIAS) for Measurement of Food Access: Indicator Guide (Washington, DC).
- Cochrane, L., Bekele, Y.W., 2018. Average crop yield (2001–2017) in Ethiopia: trends at national, regional and zonal levels. Data Brief 16, 1025–1033. <https://doi.org/10.1016/j.dib.2017.12.039>.
- Dale, V.H., Beyeler, S.C., 2001. Challenges in the development and use of ecological indicators. *Ecol. Indic.* 1, 3–10. [https://doi.org/10.1016/S1470-160X\(01\)00003-6](https://doi.org/10.1016/S1470-160X(01)00003-6).
- de Olde, E.M., Oudshoorn, F.W., Sørensen, C.A.G., Bokkers, E.A.M., de Boer, I.J.M., 2016. Assessing sustainability at farm-level: lessons learned from a comparison of tools in practice. *Ecol. Indic.* 66, 391–404. <https://doi.org/10.1016/j.ecolind.2016.01.047>.
- Dominguez-Hernandez, M.E., Zepeda-Bautista, R., Valderrama-Bravo, M.C., Dominguez-Hernandez, E., Hernandez-Aguilar, C., 2018. Sustainability assessment of traditional maize (*Zea mays* L.) agroecosystem in Sierra Norte of Puebla, Mexico. *Agroecol. Sustain. Food Syst.* 42 (4), 383–406.
- Ebert, U., Welsch, H., 2004. Meaningful environmental indices: a social choice approach. *J. Environ. Econ. Manag.* 47 (2), 270–283.
- Ellis-Jones, J., Mekonnen, K., Gebreselassie, S., Schulz, S., 2013. Challenges and Opportunities to the Intensification of Farming Systems in the Highlands of Ethiopia: Results of a Participatory Community Analysis. CIP, International Livestock Research Institute, Addis Ababa, Nairobi. <https://hdl.handle.net/10568/33789>.
- FAO, 2011. Save and Grow: A Policymaker's Guide to the Sustainable Intensification of Smallholder Crop Production. Food and agriculture organization of the United Nations, Rome (ISBN 978-92-5-106871-7).
- FAO, 2019. TAPE Tool for Agroecology Performance Evaluation 2019 – Process of Development and Guidelines for Application. Test Version (Rome).
- FAO and FH360, 2016. Minimum Dietary Diversity for Women—A Guide to Measurement. FAO, Rome.
- Fischer, G., Wittich, S., Malima, G., Sikumba, G., Lukuyu, B., Ngunga, D., Rugalabam, J., 2018. Gender and mechanization: exploring the sustainability of mechanized forage chopping in Tanzania. *J. Rural Stud.* 64, 112–122. <https://doi.org/10.1016/j.jrurstud.2018.09.012>.
- Foley, J., Ramankutty, N., Brauman, K., et al., 2011. Solutions for a cultivated planet. *Nature* 478, 337–342. <https://doi.org/10.1038/nature10452>.
- Fraval, S., Hammond, J., Bogard, J.R., et al., 2019. Food access deficiencies in sub-Saharan Africa: prevalence and implications for agricultural interventions. *Front. Sustain. Food Syst.* 3, 104. <https://doi.org/10.3389/fsufs.2019.00104>.
- Gan, X., Fernandez, I.C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., Wu, J., 2017. When to use what: methods for weighting and aggregating sustainability indicators. *Ecol. Indic.* 81, 491–502. <https://doi.org/10.1016/j.ecolind.2017.05.068>.
- Garnett, T., 2014. Three perspectives on sustainable food security: efficiency, demand restraint, food system transformation. What role for life cycle assessment? *J. Clean. Prod.* 73, 10–18. <https://doi.org/10.1016/j.jclepro.2013.07.045>.
- Garnett, T., Appleby, M.C., Balmford, A., et al., 2013. Sustainable intensification in agriculture: premises and policies. *Science* 341 (6141), 33–34. <https://doi.org/10.1126/science.1234485>.
- Gerrard, C.L., Smith, L.G., Padel, S., Pearce, B., Hitchings, R., Measures, M., Cooper, N., 2011. OCIS Public Goods Tool Development. Organic Research Centre Report. Organic Research Centre, Berkshire, UK.
- Glopan 2020, 2020. Global panel on agriculture and food systems for nutrition. In: Future Food Systems: For People, Our Planet, and Prosperity (London, UK).
- Godfray, H.C.J., 2015. The debate over sustainable intensification. *Food Sec.* 7, 199–208. <https://doi.org/10.1007/s12571-015-0424-2>.
- Godfray, H.C., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Robinson, S., Thomas, S.M., Toulman, C., 2010. Food security: the challenge of feeding 9 billion people. *Science* 327 (5967), 812–818. <https://doi.org/10.1126/science.1185383>.
- Goldstein, J.H., Caldarone, G., Duarte, T.K., Ennaanay, D., Hannahs, N., Mendoza, G., Daily, G.C., 2012. Integrating ecosystem-service trade-offs into land-use decisions. *Proc. Natl. Acad. Sci.* 109 (19), 7565–7570.
- Haileslassie, A., Craufurd, P., Thiagarajah, R., Kumar, S., Whitbread, A., Rathor, A., Blummel, M., Ericsson, P., Kakumanu, K.R., 2016. Empirical evaluation of sustainability of divergent farms in the dryland farming systems of India. *Ecol. Indic.* 60, 710–723.
- Hammond, J., et al., 2017. The Rural Household Multi-Indicator Survey (RHOMIS) for rapid characterisation of households to inform climate smart agriculture interventions: description and applications in East Africa and Central America. *Agric. Syst.* 151, 225–233.
- Hartung, C., Anokwa, Y., Brunette, W., Lerer, A., Tseng, C., Borriello, G., 2010. Open data kit: tools to build information services for developing regions. In: Proceedings of the International Conference on Information and Communication Technologies and Development, pp. 1–11. Available at: papers2://publication/uuid/ACE2FDB0-CD33-475FA750-0331358C1976.
- HLPE, 2020. Food security and nutrition: building a global narrative towards 2030. In: A Report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security (Rome).
- ILRI, 2015a. Africa RISING Ethiopia Project in Amhara Region. ILRI, Nairobi, Kenya. <https://hdl.handle.net/10568/61850>.
- ILRI, 2015b. Africa RISING Ethiopia Project in Tigray Region. ILRI, Nairobi, Kenya. <https://hdl.handle.net/10568/61851>.
- ILRI, 2015c. Africa RISING Ethiopia Project in Southern Nations, Nationalities and Peoples Region. ILRI, Nairobi, Kenya. <https://hdl.handle.net/10568/61852>.
- ILRI, 2015d. Africa RISING Ethiopia Project in Oromia Region. ILRI, Nairobi, Kenya. <https://hdl.handle.net/10568/61856>.
- Inwood, Eichler, Sarah E., et al., 2018. Assessing sustainability in agricultural landscapes: a review of approaches. *Environ. Rev.* 26 (3), 299–315.
- IPCC, 2006. 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Available at: <http://www.ipcc-nggip.iges.or.jp/public/2006gl/index.htm>.
- IPCC, 2020. Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems. P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacem, J. Malley. (ISBN 978-92-9169-154-8).
- Jambo, I.J., Groot, J.C.J., Descheemaeker, K., Bekunda, M., Titttonell, P., 2019. Motivations for the use of sustainable intensification practices among smallholder farmers in Tanzania and Malawi. *NJAS Wageningen J. Life Sci.* 89, 100306. <https://doi.org/10.1016/j.njas.2019.100306>.
- Jiao, X.Q., Zhang, H.Y., Ma, W.Q., Wang, C., Li, X.L., Zhang, F.S., 2019. Science and technology backyard: a novel approach to empower smallholder farmers for sustainable intensification of agriculture in China. *J. Integr. Agric.* 18 (8), 1657–1666. [https://doi.org/10.1016/S2095-3119\(19\)62592-X](https://doi.org/10.1016/S2095-3119(19)62592-X).
- Kanter, D.R., et al., 2016. Evaluating agricultural trade-offs in the age of sustainable development. *Agric. Syst.* 163, 73–88. <https://doi.org/10.1016/j.agsy.2016.09.010>.
- Klapwijk, C.J., van Wijk, M.T., Rosenstock, T.S., van Asten, P.J.A., Thornton, P.K., Giller, K.E., 2014. Analysis of trade-offs in agricultural systems: current status and way forward. *Curr. Opin. Environ. Sustain.* 6, 110–115. <https://doi.org/10.1016/j.cosust.2013.11.012>.
- Lee, H., 2018. Teff, a rising global crop: current status of teff production and value chain. *Open Agric.* 12, 185–193. <https://doi.org/10.2174/1874331501812010185>.
- Loos, J., Abson, D.J., Chappell, M.J., Hanspach, J., Mikulcak, F., Tichit, M., Fischer, J., 2014. Putting meaning back into “sustainable intensification”. *Front. Ecol. Environ.* 12, 356–361. <https://doi.org/10.1890/130157>.
- Lunt, Tobias, Ellis-Jones, Jim, Mekonnen, Kindu, Schulz, Steffen, Thorne, Peter, Schulte-Geldermann, Elmar, Sharma, Kalpana, 2018. Participatory community analysis: identifying and addressing challenges to Ethiopian smallholder livelihoods. *Dev. Pract.* 28 (2), 208–226. <https://doi.org/10.1080/09614524.2018.1417354>.
- Marchand, F., Debruyne, L., Triste, L., Gerrard, C., Padel, S., Lauwers, L., 2014. Key characteristics for tool choice in indicator-based sustainability assessment at farm level. *Ecol. Soc.* 19 (3), 46. <https://doi.org/10.5751/ES-06876-190346>.
- Marinus, W., Ronner, E., van de Ven, G.W.J., Kanampiu, F., Adjei-Nsiah, S., Giller, K.E., 2018. The devil is in the detail! Sustainability assessment of African smallholder farming. In: Bell, S., Morse, S. (Eds.), *Routledge Handbook of Sustainability Indicators*. Routledge, pp. 427–450. <https://doi.org/10.4324/9781315561103-28>.
- Marinus, W., Descheemaeker, K., van de Ven, G.W.J., Waswa, W., Mukalama, J., Vanlauwe, B., Giller, K.E., 2021. “That is my farm” – an integrated co-learning approach for whole-farm sustainable intensification in smallholder farming. *Agric. Syst.* 188, 103041. <https://doi.org/10.1016/j.agsy.2020.103041>.
- Miettinen, K., 2014. Survey of methods to visualize alternatives in multiple criteria decision making problems. *OR Spectr.* 36 (1), 3–37. <https://doi.org/10.1007/s00291-012-0297-0>.
- Minten, B., Taffesse, A.S., Brown, P., 2018. The Economics of Teff: Exploring Ethiopia's Biggest Cash Crop. International Food Policy Research Institute, Washington, DC.
- Mulema, A.A., Jogo, W., Damtew, E., Mekonnen, K., Thorne, P., 2019. Women farmers' participation in the agricultural research process: implications for agricultural sustainability in Ethiopia. *Int. J. Agric. Sustain.* 17 (2), 127–145. <https://doi.org/10.1080/14735903.2019.1569578>.
- Musumba, M., Grabowski, P., Palm, C., Snapp, S., 2017. Guide for the Sustainable Intensification Assessment Framework. <https://doi.org/10.13140/RG.2.2.13048.75521>.
- Mutyasira, V., Hoag, D., Pendell, D., Manning, D.T., Berhe, M., 2018. Assessing the relative sustainability of smallholder farming systems in Ethiopian highlands. *Agric. Syst.* 167, 83–91. <https://doi.org/10.1016/j.agsy.2018.08.006>.
- Njuki, J., Poole, J., Johnson, N., Baltenweck, I., Pali, P., Lokman, Z., Mburu, S., 2011. Gender, Livestock and Livelihood Indicators. ILRI, Nairobi.
- Ortiz-Crespo, B., Steinke, J., Quirós, C.F., van de Gevel, J., Daudi, H., Mgimiloko, M.G., van Etten, J., 2020. User-centred design of a digital advisory service: enhancing public agricultural extension for sustainable intensification in Tanzania. *Int. J. Agric. Sustain.* <https://doi.org/10.1080/14735903.2020.1720474>.
- Pearce, D.W., Atkinson, G.D., Dubourg, W.R., 1994. The economics of sustainable development. *Annu. Rev. Energy Environ.* 19 (1), 457–474.

- Pollesch, N., Dale, V.H., 2015. Applications of aggregation theory to sustainability assessment. *Ecol. Econ.* 114, 117–127. <https://doi.org/10.1016/j.ecolecon.2015.03.011>.
- Pound, B., Tolera, A., Matsaert, H., 2015. Report of the Internally-Commissioned External Review of the Africa RISING Project in the Ethiopian Highlands. ILRI, Nairobi, Kenya. Link to cite or share this item: <https://hdl.handle.net/10568/66567>.
- Pretty, J., 2008. Agricultural sustainability: concepts, principles & evidence. *Philos. Trans. R. Soc. Lond. B* 363, 447–466. <https://doi.org/10.1098/rstb.2007.2163>.
- Pretty, J., Benton, T.G., Bharucha, Z.P., Dicks, L.V., Flora, C.B., Godfray, H.C.J., Goulson, D., Hartley, S., Lampkin, N., Morris, C., Pierzynski, G., Prasad, V., Reganold, J., Rockstrom, J., Smith, P., et al., 2018. Global assessment of agricultural system redesign for sustainable intensification. *Nat. Sustain.* 1 (8), 441–446 (ISSN 2398-9629).
- R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Reich, J., Paul, S.S., Snapp, S.S., 2021. Highly variable performance of sustainable intensification on smallholder farms: a systematic review. *Glob. Food Secur.* 30, 100553. <https://doi.org/10.1016/j.gfs.2021.100553>.
- Rockström, J., Williams, J., Daily, G., et al., 2017. Sustainable intensification of agriculture for human prosperity and global sustainability. *Ambio* 46, 4–17. <https://doi.org/10.1007/s13280-016-0793-6>.
- Rouleau, M., Zupko, R., 2019. Agent-based modeling for bioenergy sustainability assessment. *Landsc. Urban Plan.* 188, 54–63. <https://doi.org/10.1016/j.landurbplan.2019.04.019>.
- Schreiner, M., Chen, S., 2009. Simple Poverty Scorecard Ethiopia. Available from. <http://simplepovertyscorecard.com/>.
- Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, A.K., 2012. An overview of sustainability assessment methodologies. *Ecol. Indic.* 15 (1), 281–299. <https://doi.org/10.1016/j.ecolind.2011.01.007>.
- Smith, A., Snapp, S., Chikowo, R., Thorne, P., Bekunda, M., Glover, J., 2017. Measuring sustainable intensification in smallholder agroecosystems: a review. *Glob. Food Secur.* 12, 127–138. <https://doi.org/10.1016/j.gfs.2016.11.002>.
- Snapp, S., Grabowski, P., Chikowo, R., Smith, A., Anders, E., Sirrine, D., Chimonyo, V., Bekunda, M., 2018. Maize yield and profitability tradeoffs with social, human and environmental performance: is sustainable intensification feasible? *Agric. Syst.* 162, 77–88. <https://doi.org/10.1016/j.agsy.2018.01.012>.
- Tavener, K., van Wijk, M., Fraval, S., et al., 2019. Intensifying inequality? Gendered trends in commercializing and diversifying smallholder farming systems in East Africa. *Front. Sustain. Food Syst.* 3, 10. <https://doi.org/10.3389/fsufs.2019.00010>.
- Thomson, A.M., Ellis, E.C., Grau, H.R., Kummerle, T., Meyfroidt, P., Ramankutty, N., Zeleke, G., 2019. Sustainable intensification in land systems: trade-offs, scales, and contexts. *Curr. Opin. Environ. Sustain.* 38, 37–43. <https://doi.org/10.1016/j.cosust.2019.04.011>.
- Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. *PNAS* 108 (50), 20260–20264. <https://doi.org/10.1073/pnas.1116437108>.
- Tittonell, P., Giller, K.E., 2013. When yield gaps are poverty traps: the paradigm of ecological intensification in African smallholder agriculture. *Field Crop Res.* 143, 76–90. <https://doi.org/10.1016/j.fcr.2012.10.007>.
- Valdivia, R.O., Antle, J.M., Stoorvogel, J.J., 2012. Coupling the trade-off analysis model with a market equilibrium model to analyze economic and environmental outcomes of agricultural production systems. *Agric. Syst.* 110, 17–29.
- van Loon, M.P., Deng, N., Grassini, P., Edreira, J.I.R., Wolde-meskel, E., Baijukya, F., Marrou, H., van Ittersum, M.K., 2018. Prospect for increasing grain legume crop production in East Africa. *Eur. J. Agron.* 101, 140–148. <https://doi.org/10.1016/j.eja.2018.09.004>.
- van Wijk, M., Hammond, J., Gorman, L., et al., 2020. The Rural Household Multiple Indicator Survey, data from 13,310 farm households in 21 countries. *Sci. Data* 7, 46. <https://doi.org/10.1038/s41597-020-0388-8>.
- Vang Rasmussen, L., Coolsaet, B., Martin, A., et al., 2018. Social-ecological outcomes of agricultural intensification. *Nat. Sustain.* 1, 275–282. <https://doi.org/10.1038/s41893-018-0070-8>.
- Weltin, M., Zasada, I., Piorr, A., Debolini, M., Geniaux, G., Perez, O.M., Scherer, L., Marco, L.T., Schulp, C.J.E., 2018. Conceptualising fields of action for sustainable intensification – a systematic literature review and application to regional case studies. *Agric. Ecosyst. Environ.* 257, 68–80. <https://doi.org/10.1016/j.agee.2018.01.023>.
- Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- World Bank, 2021. International Comparison Program, World Development Indicators Database. World Bank. <https://data.worldbank.org/indicator/PA.NUS.PPP> (accessed 08/07/2021).
- Yilma, Z., Emannuelle, G.B., Ameha, S., 2011. A review of the Ethiopian dairy sector. In: Fombad, Rudolf (Ed.), *Food and Agriculture Organization of the United Nations. Sub Regional Office for Eastern Africa (FAO/SFE), Addis Ababa, Ethiopia*, p. 81.