Appendix 2. Additional information on the data collection and analysis

1. Food security indicator methodology

Food security is multidimensional and has no single internationally recognized measure (Hendriks *et al*. 2016). Therefore, we used seven internationally recognized food security indicators to evaluate and compare the groups in the two study areas. The methodology to calculate these is detailed below.

The Household Dietary Diversity Score (HDDS) is a recognized measure of diet quality (Hendriks *et al*. 2016; Hirvonen, Taffesse, & Hassen, 2016; IFPRI 2006). The HDDS captured the number of food groups consumed within the previous 24 hours (FANTA 2006). The score is the sum of binary responses regarding the consumption of 12 food groups. We grouped households into by lowest dietary diversity (HDDS ≤ 3), medium dietary diversity (HDDS 4 and 5), and high dietary diversity (HDDS ≥ 6) (FAO 2006).

The Food Consumption Score (FCS) is derived from a seven-day recall similar to the HDDS (WFP 2008). The FCS is a composite score of the frequency of consumption over the previous seven days and then weighted by a coefficient (Hendriks *et al*. 2016; Leroy 2015; WFP 2008). The score is obtained as follows:

FCS = (days of staple consumption) *2 + (number of days pulses consumed) *3 + (number of days vegetables and leaves consumed) + (number of days fruit consumed) + (number of days meat/fish/eggs consumed) *4 + (number of days dairy consumed) *4 + (number of days sugar/honey consumed) * 0.5 + (number of days of oils and fats consumed) *0.5.

The results were classified as: 0–21 for poor food consumption, 21.5–35 for borderline food consumption, and above 35 for acceptable food consumption (WFP 2008).

The Women's Dietary Diversity Score (WDDS) assessed the micronutrient adequacy of the diets of women of reproductive age (15–49 years of age) (FAO and FHI, 2016). For this indicator, we could only use the data for female-headed households and assume that the responses to the questions on consumption by the household head reflected women's dietary patterns. The score was also derived from the 24-hour recall food consumption data, but we reclassified the responses according to nine food groups based on nutritional importance (Chagomoka *et al*. 2016; FAO 2011; Kennedy 2010; Leroy 2015). The WDDS was classified into three categories according to (Chagomoka *et al*. 2017).

The Months of Adequate Household Food Provisioning (MAHFP), measures household food access over a year (Africare 2007; Bilinsky and Swindale 2010). The score was the sum of the months of adequate provision (Bilinsky and Swindale 2010). The households were classified into three categories as indicated by Africare (2007).

The Coping Strategies Index (CSI) indirectly measures food security by asking questions related to food consumption behaviour (Hendriks *et al*. 2016; Leroy 2015; Maxwell and Caldwell 2008). The CSI was calculated by multiplying the frequency and severity of behaviours that households engaged in to mitigate food shortages from a seven-day recall period following Maxwell and Caldwell's (2008) methodology:

$$CSI = (frequency\,CS\,1 * severity\,CS\,1) + (frequency\,CS\,2 * severity\,CS\,2) + \ldots + (frequency\,CS\,10 * severity\,CS\,10)$$

An asset ownership indicator was used as a proxy for household resilience (ability to cope with shocks) (Swift 2006). Low asset levels increase vulnerability to poverty and hunger (food insecurity) (Chambers 2006; Maxwell and Smith 1992). We used a simple sum of household assets classes. The sum does not reflect the value of assets (Browne *et al*. 2014; Hendriks *et al*. 2016).

A modified Consolidated Approach for Reporting Indicators of Food Security (CARI) console was also used for comparative analysis (WFP 2012). The CARI combines food security indicators (current status and coping capacity) into a summary called the Food Security Index (FSI), representing the overall food security status (Butaumocho and Chitiyo 2017). The CARI used a combination of three food security indicators (i.e. food consumption score, food expenditure shares, and livelihood coping strategy). Due to a lack of livelihood coping strategy-related indicators in the database, we converted the data regarding the application of the more serious consumption-based coping strategies (i.e. adult hunger, child hunger, and eating fewer meals) into a livelihood coping strategy score. Households were classified into four groups: food secure, marginally food secure, moderately food insecure, and severely food insecure) (WFP 2014).

A chi-square test was used to check the significant difference between the household groups in the two case studies and the four household groups. Spearman's correlation was used to examine the non-parametric relation between food security indicators (HDDS, FCS, MAHFP, CSI, and Asset).

2. Socio-economic and food security impacts

On socio-economic impacts and food-security impacts in Kenya, only aggregated data for the cases KE1–5 were available; we used this data as "best available data" for these LAI cases. Food security scores for Kenya were calculated without case KE6, because we were not able to include the KE6 case into formal analysis due to missing data on the indicators from other work packages.

3. Environmental impact scoring (indicators ENV1-7 and ENV 11–17)

We used 14 indicators to assess environmental impacts combining participatory and expert-based assessments of environmental impacts. Half of them (indicators ENV1–7) are based on semi-structured interviews that measure households' perceptions of chemical exposure, deforestation, overabstraction of water, water pollution with chemicals and effluents, air pollution with chemicals/respiratory problems, increase in pests and occupation of water sources. For each indicator, we quantified the share of households perceiving this environmental impact. We classified this data into very low impacts (0% of households perceive this environmental impact); low (1-33%); medium (34-66%); and high impacts (67-100%).

The other half of our data on environmental impacts (indicators ENV11–17) are based on interviews with LAI and small-scale farmers, corresponding with lifecycle assessments and expert assessments. Environmental impact scoring for the different cropping systems was based on a per surface area (one hectare) basis. Life cycle assessment (LCA) metrics for global warming potential, eutrophication potential, acidification potential, non-renewable energy consumption, and water consumption were calculated for specific case studies (Da Silva, 2018) and used to inform the scoring. Extent of pesticide consumption and soil degradation in the form of decline of soil organic matter and nutrient mining were also considered (Ottinger, 2018; Da Silva, 2018). Table A2 was further used to guide the selection of an impact score based on intensity of resource use. It is also acknowledged that some degree of subjectivity was needed to assign impacts scores to different case studies due to lack of data.

For water consumption, systems solely dependent on rainfall were scored as having “very low” or “low” impact. For Kenya, flower production in greenhouses was difficult to score. While most of these systems may operate under open fertigation thus scoring “high” impact, one must also consider that production is occurring in a more controlled environment with lower atmospheric water demand and there may be rainwater harvesting from the roof of the greenhouses for irrigation purposes. It is also possible that any over-irrigation may be returned to the system and available to other users downstream (non-consumptive, recoverable fraction). Soil degradation is an equally challenging category to score for greenhouse production. While the soil under the greenhouses is protected from erosion and crusting due to the overhead cover and micro-irrigation potentially leading to reduced impact, a large amount of agrochemicals are applied to these soils, and it was also indicated that these soil needs to be replaced every nine or ten years.

Pesticide use on its own does not represent direct impact, but it has been reported that less than 0.1% reaches the intended pest (Pimentel and Levitan 1986), so the extent of use is applied here to score the potential impact. Acidification potential is linked to the amount of energy used in the form of agro-chemical synthesis, transport and application, as well as more direct on-farm energy consumption (electricity, diesel).

An alternative to assessing the impact on a surface-area basis could have been per unit production. In many cases while the impact of an LAI is relatively higher per surface area than for a small-scale farmer, when considered per unit production the relative impact of the LAI would be lower or even more favourable compared to SSF production.

Table B1. Criteria used in addition to life-cycle assessment metrics used to score environmental impact for different case studies

Impact category	1 - Very low impact	2 - Low impact	3 - Moderate impact	4 - High impact
Pesticide use	No use of pesticides	Low use of pesticides	Moderate use of pesticides	High use of pesticides
Eutrophication potential	No use of fertilisers	Low use of fertilisers	Moderate use of fertilisers	High use of fertilisers
Acidification potential	No use of agro-chemicals and/or mechanisation	Low use of agro-chemicals and/or mechanisation	Moderate use of agro-chemicals and/or mechanisation	High use of agro-chemicals and/or mechanisation
Global warming potential	No use of agro-chemicals and/or mechanisation	Low use of agro-chemicals and/or mechanisation	Moderate use of agro-chemicals and/or mechanisation	High use of agro-chemicals and/or mechanisation
Non-renewable energy conumption	High resource use efficiency	Moderate resource use efficiency	Low resource use efficiency	Very low resource use efficiency
Water consumption (blue and green water)	Very low water consumption	Low water consumption	Moderate water consumption	High water consumption
Soil degradation	High use of soil conservation measures	Moderate use of soil conservation measures	Low use of soil conservation measures	No use of soil conservation measures

The **standards for LCA methodology** were set by the International Organization for Standardization (ISO) and were defined in ISO 14040 (ISO 2006). This methodology was applied in this study to assess a range of Environmental Impacts (EIs), using LCA methodology to quantify potential off-site environmental impact indicators, namely: Eutrophication Potential (EP), Acidification Potential (AP), Global Warming Potential (GWP), the Water Footprint (WF), and Non-Renewable Energy (NRE) consumption. Further, the APSIM model calibrated with local soil and weather data was used to quantify on-site soil degradation (soil organic carbon (C) and total soil nitrogen (N) depletion) and investigate yield gaps for each system.

4. Land use changes

For the truth table, we transformed percentage values of indicators into four classes (0%, 1–33%, 34–66%, 67–100%). We treated the land cover and land use change (LCLUC) data in the following ways: (1) Distinguish on-site (“LAI”) and off-site LCLUC (“Doughnut”). (2) Calculate net change per LCLU class (ha). (3) Identify and flag the two most increasing and the two most decreasing LCLU classes per LAI case and per region. Comparison with relative LCLUC (%) and with stable LCLU classes (ha) to check whether changes are large or small. (4) Assign “1” to the two most growing and most diminishing LUCs per LAI case: on-site and off-site.

Off-site LCLUC in the cases MO1–3 cannot be assigned to a specific LAI because these LAIs are very close to each other. Therefore, we assigned the same LCLUC values to each of the three LAI cases MO1–3.

5. National governance systems

We used the data on national governance systems and regional social-ecological contexts for each LAI within that context.

Literature cited

Africare. 2007. *Guidance: How to measure Months of Adequate Household Food Provisioning (MAHFP) based on participatory rural appraisals in food security interventions*. Africare Food Security Review, No. 1. Africare, Washington DC, USA.

Bilinsky, P., and A. Swindale. 2010. *Months of adequate household food provisioning (MAHFP) for measurement of household food access: Indicator guide*. Version 4. Washington DC, USA: Food and Nutrition Technical Assistance II Project (FANTA).

Browne, M., G. F. Ortmann, and S. L. Hendriks. 2014. Developing a resilience indicator for food security monitoring and evaluation. *Agrekon* 53(2):25:46.

Butaumocho, B., and P. T. Chitiyo. 2017. A comparative analysis of household food security measure in rural Zimbabwe. *International Journal of Food and Agricultural Economics* 5(2):41-58.

Chagomoka, T., A. Drescher, R. Glaser, B. Marschner, J Schlesinger, and G. Nyandoro. 2017. Contribution of urban and periurban agriculture to household food and nutrition security along the urban–rural continuum in Ouagadougou, Burkina Faso. *Renewable Agriculture and Food Systems* 32(1):5-20.

FANTA (Food and Nutrition Technical Assistance Project). 2006. *Household Dietary Diversity Score (HDDS) for measurement of household food access: Indicator guide (Version 2).* FANTA, Washington DC, USA.

FAO (Food and Agricultural Organization of the United Nations). 2006. *Food Security. Policy Brief No. 2. June.* FAO, Rome, Italy.

FAO, FHI 360. 2016. *Minimum dietary diversity for women: A guide for measurement.* FAO, Rome, Italy. 2016.

Hendriks, S. L., C. van der Merwe, M. S. Ngidi, C. Manyamba, M. Mbele, A. M. McIntyre, E. Mkandawire, Q. N. Molefe, M. Q. Mphephu, and L. Ngwane. 2016. What are we measuring? Comparison of household food security indicators in the Eastern Cape Province, South Africa. *Ecology of food and nutrition* 55(2):141-162.

Hirvonen, K., A. S. Taffesse, I. W. Hassen. 2016. Seasonality and household diets in Ethiopia. Public Health Nutrition 19(10):1723-1730.

IFPRI (International Food Policy Research Institute). 2006. *Review and validation of dietary diversity, food frequency and other proxy indicators of household food security*. Report submitted to the World Food Programme, Rome, Italy.

ISO (International Organization for Standardization). 2006. Environmental management - Life cycle assessment – Principles and framework. ISO 14040:2006. ISO, Geneva, Switzerland.

Kennedy, G., M. Razes, T. Ballard, and M. C. Dop. 2010. *Measurement of dietary diversity for monitoring the impact of food based approaches.* In: International symposium on food and nutrition security: food-based approaches for improving diets and raising levels of nutrition. FAO, Rome, Italy.

Leroy, J. L., M. Ruel, E. A. Frongillo, J. Harris, and T. J. Ballard. 2015. Measuring the food access dimension of food security: a critical review and mapping of indicators. *Food and Nutrition Bulletin* 36(2):167-195.

Maxwell, D., and R. Caldwell. 2008. *The Coping Strategies Index: Field methods manual.* 2nd edition. Cooperative for Assistance and Relief Everywhere (CARE), USA.

Maxwell, S., and M. Smith. 1992. Household food security: a conceptual review. In: Maxwell, S., and T. Frankenberger (eds.) *Household food security: concepts, indicators, measurements*. International Fund for Agricultural Development (IFAD) and United Nations International Children's Emergency Fund (UNICEF). Rome and New York, pp. 1-72.

Swift, J. 2006. Why are rural people vulnerable to famine? *IDS Bulletin* 37(4):41-49.

WFP (World Food Programme). 2012. *Guidance note: Calculation of household food security outcome indicators* (WFP Vulnerability Analysis and Mapping Unit, Afghanistan December 2012). Available from: https://www.humanitarianresponse.info/ru/system/files/documents/files/guidance_note_-_calculation_of_fcs_rcsi_hhs_and_dd.docx; Accessed May 02, 2020.

WFP (World Food Programme). 2008. *Vulnerability analysis and mapping. Food consumption analysis: Calculation and use of the food consumption score in food security analysis*. WFP, Rome, Italy.