

<https://doi.org/10.1038/s44264-024-00039-9>

Yield gap decomposition: quantifying factors limiting soybean yield in Southern Africa

Check for updates

John Okoth Omondi¹ , Mazvita Sheila Chiduwa², Stephen Kyei-Boahen³, Patricia Masikati⁴ & Isaiah Nyagumbo^{2,5}

Soybean production in Sub-Saharan Africa (SSA) is increasing as its demand for food, feed, cash, and soil fertility improvement soars. Yet, the difference between the smallholder farmers' yield and either the attainable or the potential is large. Here, we assessed the contribution of various crop management practices to yield gap, and the major factors limiting soybean yield increase per unit area. This study showed that besides soil nutrients and plant nutrition, soybean variety is the most limiting factor in Malawi and Zambia, whereas, in Mozambique, seed rate is significant. Overall, in the Southern Africa region (Malawi, Zambia, and Mozambique) the major soybean yield gap contributors are: variety (63.9%), seed rate (49.7%), and disease damage (36.3%), especially soybean rust. An indication that through yield gap decomposition, interventions could be prioritized to target the most yield-limiting factors with the minimum resources available to smallholder farmers and immensely narrow the yield gap.

Soybean (*Glycine max (L.) Merril*) is a legume crop that improves soil fertility¹, human and animal nutrition², and income³. The demand for soybeans continues to rise globally, while its production in Southern Africa is low, only contributing to 1% of the global output⁴. From this, commercial production dominates, and South Africa produces about 65% of the total regional produce, followed by Zambia, Malawi, and Zimbabwe⁵. Contribution by smallholder farmers (owning farms that are less or equal to 2 hectares⁶) is limited. This is attributed mostly to marginal profits due to low yields, in both quality and quantity, resulting from poor agronomic management practices, poor or low input use, poor soil fertility, inappropriate variety choices specific to agro-ecological zones, and erratic rainfall patterns and amounts⁷. Moreover, most governments in the region focus on staple food crops like maize^{8,9}, leaving minimal support for the soybean value chain. Nonetheless, there is an increase in hectarage under soybean production.

Soybean production in Sub-Saharan Africa (SSA) is increasing as a result of hectarage under production^{5,10}, but not yield increase per hectare. Undeniably, the increase in hectarage under soybean will continue¹¹ and as observed from previous trends (from 2010 to 2020) (Omondi et al.⁵). This is driven by improved market prices¹² as demand for soybeans as food and feed, especially in the poultry industry, rise¹³; diversification from maize^{14,15} and particularly to enhance climate change adaptation and resilience¹⁶ and

improve smallholder farmers' nutrition¹⁷; improve soil fertility through biological nitrogen fixation¹⁸; break pest and disease cycles—pest control in integrated pest management such as fall armyworm and stemborer^{19,20} among others. Such land expansion for more production will gradually be unsustainable as the population grows; the land size per household reduces^{21,22} and hence yield per unit area requires improvement²³. This yield increase per unit area could be realized by understanding the factors that contribute to the yield gap for specific agro-ecologies or individual farmers.

In the context of rainfed crops, yield gap is defined as the difference between the water-limited yield (Y_w) and the actual yield (Y_a) observed in farmers' fields²⁴. The water-limited yield (Y_w) is defined as the maximum yield that can be obtained under rain-fed conditions in a specific biophysical environment without nutrient limitations or other yield reduction factors such as pests, diseases, weeds etcetera²⁵. Therefore, yield gap decomposition enables portioning of each yield contributing factor⁷, for example, soil nutrient, seed rate, variety, weeds, diseases and pests etcetera to their respective yield gap value²⁶. It thus strengthens targeting of specific interventions to specific agro-ecologies or farmers depending on their capability to implement such interventions or innovations i.e., if soybean variety is the most limiting, then it is given the primary priority, and more resources are deployed to narrow the yield gap.

¹International Institute of Tropical Agriculture (IITA), Chitedze Research Station, Lilongwe, Malawi. ²International Maize and Wheat Improvement Center (CIMMYT), Chitedze Research Station, Lilongwe, Malawi. ³International Institute of Tropical Agriculture (IITA), Av. FPLM via Corrane, Nampula, Mozambique. ⁴World Agroforestry (ICRAF), St. Eugene Office Park, 39 P Lake Road Kabulonga, Lusaka, Zambia. ⁵International Maize and Wheat Improvement Center (CIMMYT), P.O. Box MP163 Mazowe Road, Mount Pleasant, Harare, Zimbabwe. e-mail: okoth05@gmail.com

Omondi et al.⁵ reported that the average soybean yield for Malawi in 2023 based on data from FAOSTAT was 39.8% lower while those of Zambia and Mozambique were 6.1% and 25.9% higher, respectively than Africa's average yield in 2023. This indicates that the causes of yield gaps are country-specific and could further narrow to agro-ecologies, households, and individual farms. They also noted that in comparison to the global average yield, the yield gap widened in all three countries; the average soybean yield in the world was 195.2%, 99.0%, and 67.8% higher than in Malawi, Zambia, and Mozambique, respectively. Accordingly, they concluded that globally, some countries were narrowing their yield gaps and closer inching to the potential yield, while most countries in Sub-Saharan Africa lagged. They further compared FAOSTAT soybean yield in 2020 with the attainable soybean yield under breeding trials at the International Institute of Tropical (IITA) in those three countries and observed wide yield gaps for Malawi, Zambia, and Mozambique, respectively. This was even wider than the comparison with either Africa's or the global average yield. Therefore, decomposing the yield gap i.e., identifying the causes of yield gaps in these countries and prioritizing interventions to narrow the gaps is necessary. Thus, this study assessed the contribution of various crop management practices and inputs to yield gap, the actual farmer-attainable soybean yield for each country, and the major factors limiting soybean productivity.

Results

Fertilizer application (quantity and type) to soybean plants was also a question posed to the farmers, only <5% responded positively, stating that they applied fertilizers. Owing to this low response, boundary line analysis of fertilizer effect on soybean yield wasn't conducted in the three countries.

Despite the availability of some fertilizer blends for soybean in the market, for example, soya mix A or B (7-20-13 + S+Zn or 5-20-20 + S +Zn), many farmers in Malawi, Zambia, and Mozambique (approximately 95% of soybean farmers surveyed) do not apply fertilizer to legumes, including soybean.

Seed rate of soybean

The soybean seed rate in Malawi ranged from 6.2 to 152.8 kg ha⁻¹, however, the highest actual farmer yield (4882.7 kg ha⁻¹) was obtained at 85.0 kg ha⁻¹ (Fig. 1a, Table 1), and the yield gap was 27.3% (Fig. 2a, Table 1). The yield gap from 6.3% farms and upto 602.9 kg ha⁻¹ grain yield couldn't be explained by seed rate (Table 1) i.e., the data points that fell below the 1:1 line in the observed versus predicted yield graph (Fig. 1b). In Zambia, the lowest actual farmers' seed rate was 19.6 kg ha⁻¹ and the highest was 165.2 kg ha⁻¹ producing 1049.5 kg ha⁻¹ and 2500 kg ha⁻¹ soybean grain yield, respectively (Fig. 1c, Table 1). This highest seed rate also produced the highest grain yield. The resulting yield gap was 58.0% (Fig. 2b), and the seed rate could explain 93.7% of this (Fig. 1d, Table 1). The actual farmers' seed rate in Mozambique ranged from 14.2 to 293.5 kg ha⁻¹ leading to 469.5 and 1299.0 kg ha⁻¹ soybean grain yield, respectively (Fig. 1e, Table 1). However, 31.9 kg ha⁻¹ produced the highest actual farmer grain yield of 1737.0 kg ha⁻¹ in Mozambique. The yield gap between the highest actual farmer's yield and the lowest was 63.8% (Fig. 1c), and in 8.9% of the farms, this yield gap could not be explained by the seed rate (Fig. 2f, Table 1). In comparing the other two countries with Malawi, which had the highest actual farmer yield, a 94.4% seed increase in Zambia still led to lower grain yield by 95.5%. Conversely, in Mozambique, the seed rate at which the highest grain was produced was lower than in Malawi (166.5% less), yet, soybean grain yield in Malawi was higher by 181.1%. Overall, the highest soybean yield gap resulting from seed rate was from Mozambique, followed by Zambia and then Malawi (Fig. 2 and Table 1). Additionally, yield gap from many farms in Mozambique compared to the other countries could not be explained by the seed rate.

Soybean variety

The soybean varieties planted by the farmers in the three countries included: Tikolore, Makwacha, Nasoko, and Serenade in Malawi; Kafue (also called Tikolore in Malawi), Dina, Kamulanga (a local variety) in Zambia; Wamini (same variety as Tikolore and Kafue), Safari, and Serenade in Mozambique (Fig. 3). In Malawi, the highest actual farmers' yield influenced by soybean variety was produced by Tikolore (4882.7 kg ha⁻¹), whereas the lowest was

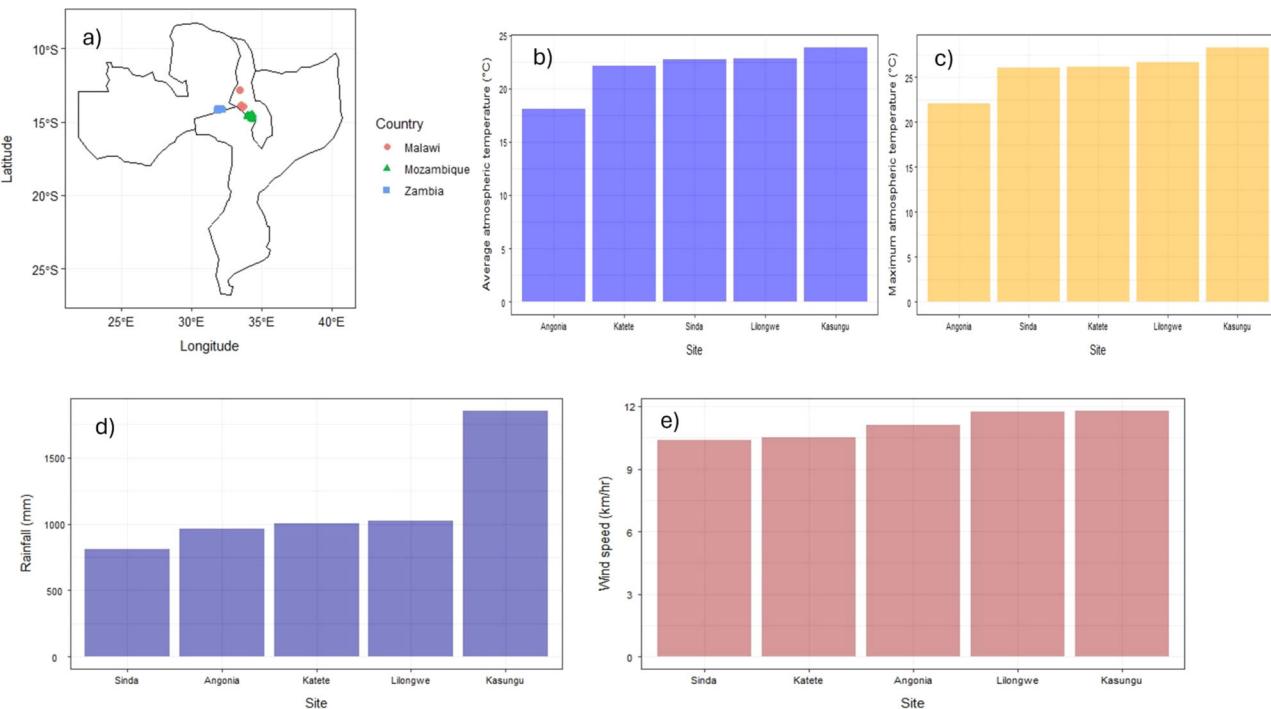


Fig. 1 | The distribution of the households and the climatic characteristics in the study area. a Study area showing the distribution of households in Malawi, Zambia, and Mozambique; **b, c** the average and maximum atmospheric temperature,

respectively; **d** rainfall amount and **e** wind speed during the period during which the farmers reported soybean being in the field in various site.

Table 1 | Percent yield gap and the unexplained yield gap by various management or inputs in Malawi, Zambia, and Mozambique

Country	Yield gap (%)	% of farms with yield gap unexplained by the management/input	Least and highest yield gap unexplained by the management/input (kg/ha)
Management/input			
Seed rate (kg ha ⁻¹)			
Malawi	27.3	6.3	14.2–602.9
Zambia	58.0	6.3	42.6–595.1
Mozambique	63.8	8.9	11.3–457.4
Soybean variety			
Malawi	64.9	0.0	0.0
Zambia	66.7	2.3	300.6–446.9
Mozambique	60.1	3.2	60.8–274.6
Disease damage			
Malawi	62.1	4.6	110.5–542.2
Zambia	22.5	1.7	46.1–254.3
Mozambique	24.3	1.9	9.7–188.1
Pest damage			
Malawi	38.4	3.4	59.1–467.9
Zambia	45.7	1.1	55.8–240.7
Mozambique	20.6	2.4	15.3–110.0
Weed management			
Malawi	34.9	4.6	17.1–558.9
Zambia	5.2	1.1	15.3
Mozambique	4.4	0.6	43.7

produced by Nasoko (1710.4 kg ha⁻¹) (Fig. 3a)—a yield gap of 64.9% (Fig. 6a and Table 1). This yield gap could be fully explained by the choice of soybean variety planted (Fig. 3b, Table 1). In Zambia, as in Malawi, Kafue variety (is Tikolore in Malawi) outperformed the rest, producing 2500 kg ha⁻¹ against the lowest yield of 833.3 kg ha⁻¹ from Kamulanga (Fig. 3c). This led to a 66.7% yield gap (Fig. 3b and Table 1), a yield gap that in 97.7% of the farms surveyed was attributed to the choice of soybean variety planted and 446.9 kg ha⁻¹ grain yield could not be explained (Fig. 3d and Table 1). Interestingly, the same variety in both countries (Tikolore in Malawi or Kafue in Zambia) produced lower yields in Zambia compared to Malawi. Serenade produced higher yields (1660.4 kg ha⁻¹) in Mozambique compared to all the other varieties, and Safari's yield was the lowest (662.9 kg ha⁻¹) (Fig. 3e) leading to a yield gap of 60.1% (Fig. 2c and Table 1). This yield gap could not be explained by the choice of soybean variety in 3.2% of the farms surveyed (Fig. 3f, Table 1). Altogether, the yield gap due to soybean variety was highest in Zambia, followed by Malawi, and then Mozambique (Table 1). Moreover, the yield gap that could not be attributed to the choice of soybean variety planted was also higher in Zambia compared to the other countries (Table 1).

Disease damage

Disease damage was evaluated in the three countries using the scores: no sign of disease damage as '1', '2' as limited signs of disease damage, '3' as moderate signs of disease damage, and '4' as widespread signs of disease damage. In Malawi, the soybean farms with no signs of disease produced the highest yields (4848.3 kg ha⁻¹), whereas the lowest was from farms with widespread signs of disease (1837.8 kg ha⁻¹) (Fig. 4a) leading to a yield gap of 62.1% (Fig. 2a). Disease damage could not explain this yield gap in 4.6% of the farms and a grain yield of up to 542.2 kg ha⁻¹ (Fig. 4b, Table 1). The highest soybean yield when there was no disease damage in Zambia was 2500 kg ha⁻¹, while the farms with moderate signs of disease produced 1938.3 kg ha⁻¹ (Fig. 4c). This caused a yield gap of 22.5% (Fig. 2b and Table 1).

A yield gap that disease damage could explain well in 98.3% of the farms (Fig. 4d, Table 1). Similar to Malawi, farms with no signs of disease in Mozambique produced more soybean (1737.0 kg ha⁻¹) than those with widespread disease (1314.9 kg ha⁻¹) (Fig. 4d) leading to a yield gap of 24.3% (Fig. 2c and Table 1). However, in 1.9% of the farms, and up to 188.1 kg ha⁻¹ of soybean grain of the yield gap could not be associated with disease damage (Fig. 4e). Overall, the yield gap attributed to disease damage, the number of farms and grain yield that could not be explained by disease damage was highest in Malawi (Fig. 2, Table 1).

Pest damage

In the three countries, pest damage was scored using a scale of '1' as no sign of pest damage, '2' as limited signs of pest damage, '3' as moderate signs of pest damage, and '4' represented widespread signs of pest damage. Like disease damage, yield after widespread pest damage in Malawi was 3005.4 kg ha⁻¹ compared to lack of pest damage (4848.3 kg ha⁻¹) (Fig. 5a) resulting in a yield gap of 38.4% (Fig. 2a and Table 1). This yield gap could not be explained by pest damage in 3.4% of the farms, similarly, the grain yield of up to 467.9 kg ha⁻¹ (Fig. 5b, Table 1). In Zambia, soybean yields were highest in fields where there were no signs of pest damage and lowest where the pest damage was widespread (Fig. 5c). This led to a yield gap of 45.7% (Fig. 2b, Table 1) which could be explained by pest damage in 98.9% of the farms surveyed (Fig. 5d, Table 1). Maintaining pest-free soybean fields in Mozambique led to a higher yield of 1737.0 kg ha⁻¹ in comparison to the lowest yields when the pest damage was widespread (Fig. 5e). The yield gap attributed to pest damage was 20.6% (Fig. 2c and Table 1), and those from 2.4% of the farms could not be explained by pest damage (Fig. 5f, Table 1). In general, pest damage led to higher yield reduction in Zambia, and yield gap from more farms could not be explained by pest damage (Fig. 2, Table 1).

Weed management

Weed management impact in the three countries was scored on a scale of '1' representing well managed (no weed infestation), '2' being moderately managed (some weeds observed), and '3' as poorly managed (>half the field is infested with weeds). In the fields where there was no weed infestation in Malawi, soybean yield was highest compared (4882.7 kg ha⁻¹) (Fig. 6a) to the lowest in highly weed-infested fields. This led to a yield gap of 34.9% (Fig. 2a and Table 1). This yield gap and up to 558.9 kg ha⁻¹ could not be explained by weed infestation in 4.6% of the farms surveyed (Fig. 6b, Table 1). In Zambia and Mozambique, as in Malawi, well-weeded fields produced more soybean yield (Fig. 6), although the yield gaps due to weed infestation were negligible at 5.2% and 4.4% in Zambia (Fig. 2b and Table 1) and Mozambique (Fig. 7c and Table 1), respectively. Generally, weed infestation led to a higher yield gap in Malawi (Fig. 2), although, compared to Zambia and Mozambique yield gap from many farms could not be explained by weed infestation and more farms (Table 1).

Discussion

Reaching potential yield is always hindered by various factors^{5,7,27}. Identifying the most yield-limiting factors could lead to prioritization of interventions and targeted approaches to increase yields from actual to attainable to potential per unit area. Through yield gap decomposition using boundary line analysis, this study has identified the limiting factors for soybean yield increase in Malawi, Zambia, and Mozambique. Furthermore, when using boundary line analysis it is possible to check if one of the factors affecting yield in a given location can explain the identified yield gap²⁸—this has been done in this study. This not only boosts prioritization of interventions but also enhances targeting.

Soybean variety is the most yield-limiting factor in both Malawi and Zambia and in Malawi it largely explains the yield gap observed. Interestingly, the popular soybean variety that performs better in both countries is the same (Tikolore in Malawi or Kafue in Zambia or Wamini in Mozambique also called TGx 1740-2 F)—an improved promiscuous variety^{29,30}. Abate, (2012) observed that improved varieties (e.g., TGx 1740-2 F in Malawi) had a yield advantage of 38% over local check, and the results,

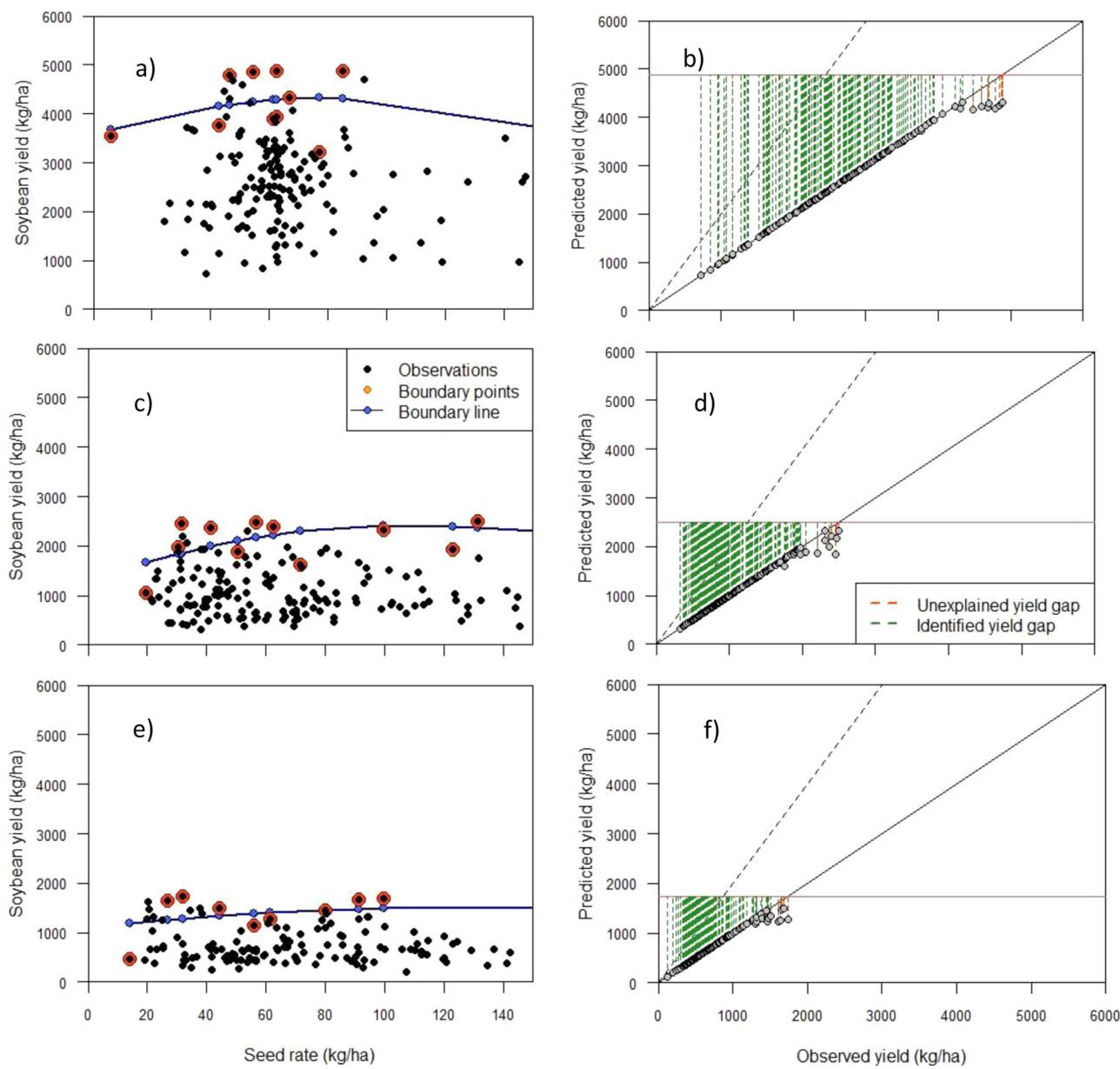


Fig. 2 | Boundary line fitted for soybean yield against seed rate, and yield gap related to the seed rate. a, c, e Boundary lines and b, d, f yield gaps in Malawi, Zambia, and Mozambique, respectively. The 1:1 represents when the actual yield is

equal to the predicted yield. The points that fall below the 1:1 line indicate yield gap that cannot be explained by the seed rate.

especially from Zambia attest to this, since the local variety (Kamulanga) performed dismally. In as much as the improved variety that performed better in both Zambia and Malawi were similar, there yield gaps with the least performing varieties were variable. The difference in varietal yields in both environments is mostly due to the differences in agro-ecologies—at Katete and Sinda in Zambia, the rainfall was 1004.5 mm and 811.2 mm, respectively whereas at Kasungu and Lilongwe, the rainfall was 1854.4 mm and 1022.5 mm, respectively during the growth period (Fig. 7). This high amount of rainfall in Malawi sites provided the required soil moisture to drive the observed yields since soybean's optimal water requirement is between 400 and 700 mm under rainfed³¹. Additionally, the soils in Kasungu could be rich in phosphorus compared to Katete and Sinda, and P is key for soybean yield improvement^{32,33}. Despite better soil water and nutrient conditions in Malawi for the soybean variety, Tikolore, the least performing variety that led to the higher yield gap probably had lower nutrient conversion efficiency. Indeed³⁴, confirms that varieties with low nutrient conversion efficiency can have dismal performance even under suitable nutrient

conditions. Disease damage is the second most limiting factor in Malawi, even though it could not explain up to 542.2 kg ha⁻¹ grain yield of the yield gap. Disease damage is the second most limiting factor in Malawi because in 2022/2023 season when the survey was conducted there was a wide outbreak of soybean rust in most of the Southern African countries^{4,35}. It was devastating in Malawi³⁵ as most farmers were unable to control the disease in time mainly because they assumed the crop was approaching physiological maturity (feedback during the survey). Certainly, they had poor knowledge to identify soybean rust symptoms, cemented by Murithi et al.⁴. Indeed, soybean rust symptoms can easily be confused with brown spots or bacteria pustules³⁶, and over time with physiological maturity, and hence it requires seasoned skills and knowledge to identify. Despite this, there is a general recommendation in the region to soybean farmers to always apply fungicide at the beginning of flowering, and then at two to three weeks intervals depending on its severity in the area³⁷. A recommendation that perhaps was observed more by farmers in Zambia, where the disease damage was minimal, compared to those in Malawi because of large-scale production in

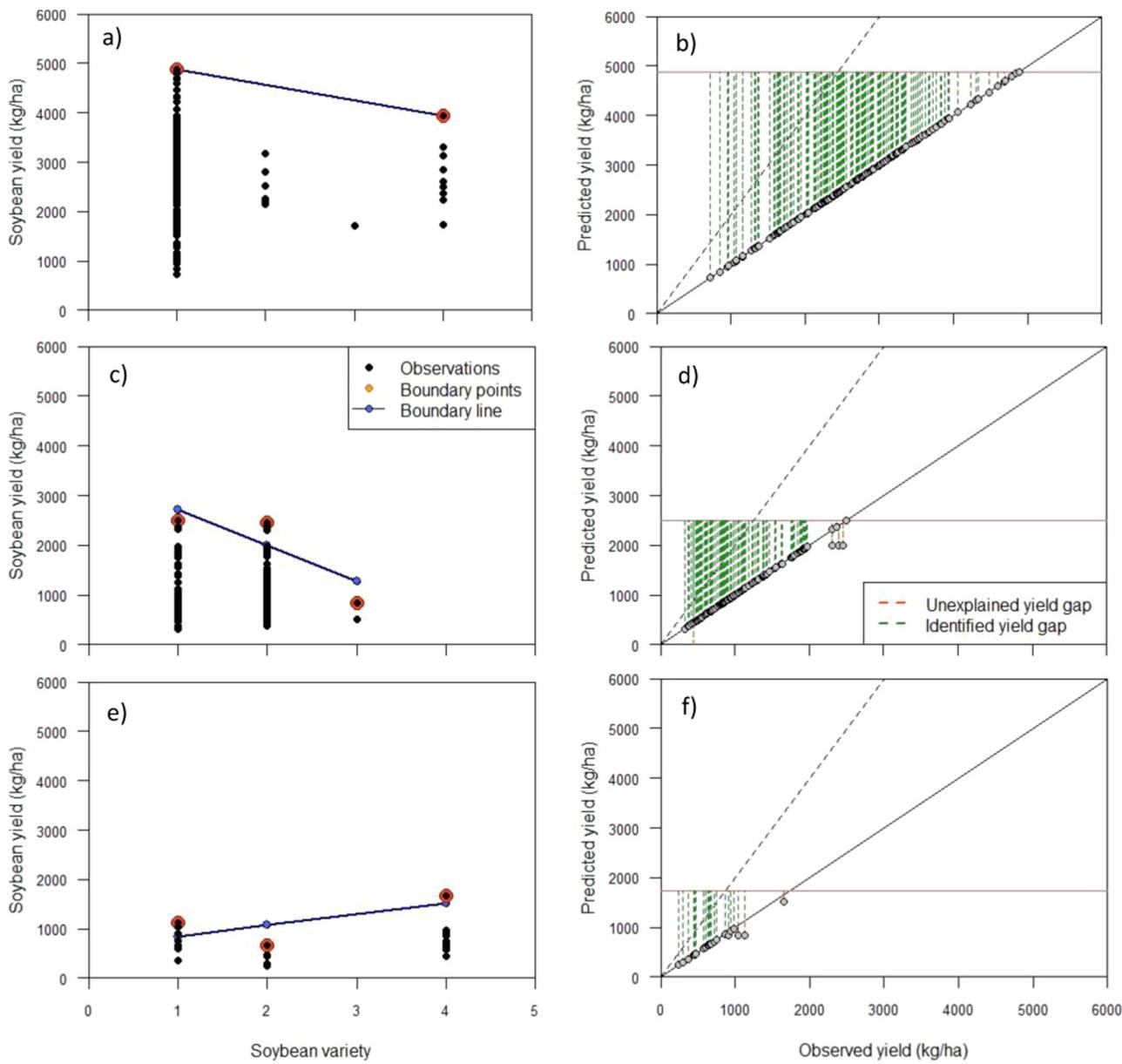


Fig. 3 | Boundary line fitted for soybean yield versus the variety planted, and yield gap driven from soybean variety. a, b represent boundary line and yield gap, respectively, in Malawi; similarly, c, d in Zambia; also, e, f in Mozambique. The names of the soybean varieties Malawi: 1—Tikolore, 2—Makwacha, 3—Nasoko, 4—Serenade; for Zambia: 1—Kafue (also called Tikolore in Malawi), 2—Dina, 3—

Kamulanga (a local variety); for Mozambique: 1—Wamini, 2—Safari, 4—Serenade. The 1:1 represents when the actual yield is equal to the predicted yield. The points that fall below the 1:1 line indicate yield gap that cannot be explained by the soybean variety planted.

Zambia⁵ hence more investment in production such as mechanization³⁸ or application of chemicals³⁹. Besides, poor skills to identify soybean rust, and lack of early control⁴⁰, high wind current⁴¹, high rainfall amount⁴², and humidity⁴³ probably contributed to vast spread and attacks of the rust in Malawi. This was the case in Malawi compared to other agro-ecologies (Zambia and Mozambique), as the wind speed in Kasungu and Lilongwe were above the other sites (Fig. 7) which encouraged dispersion of the rust spores⁴¹, coupled with high rainfall amounts (as in Fig. 7), improved the attachment and development of the spores³⁶. Moreover, the high yield gap under weed management indicates that weed pressure in Malawi was high, further increasing the host availability for rust—serving as an alternative host⁴⁴, and increasing competition for resources such as nutrients⁴⁵, thus reducing the crop vigour. It is only in Zambia where the pest damage led higher yield gap than disease damage, possibly due to the variety mostly grown by the farmers (Kafue) wasn't susceptible to pests as the local variety

(Kamulanga). Additionally, most Zambian soybean farmers sometimes intercrop soybean with sunflower, thus offering more food to pests, like *Dectes texanus*⁴⁶ and *Chrysodeixis includens*⁴⁷, increasing their fecundity and population.

While in both Malawi and Zambia, soybean variety contributed to the highest yield gap, seed rate was the highest contributor in Mozambique despite it not being to explain up to 457.4 kg ha⁻¹ grain yield of the yield gap. Remarkably, the results indicated that in Mozambique, the minimum and maximum seed rates were higher than the other two countries, an indication that a higher seed rate could lead to diminishing returns i.e., higher plant population, thus competition for resources such as light and nutrients^{48–50}. Undeniably, Silva et al.⁵¹ observed that a higher plant population increases competition for resources hence reducing yields. Strengthening this is the planting arrangement and spacing of 50 cm by 10 cm in Mozambique⁵² compared to the other countries, for example, in Malawi farmers prepare

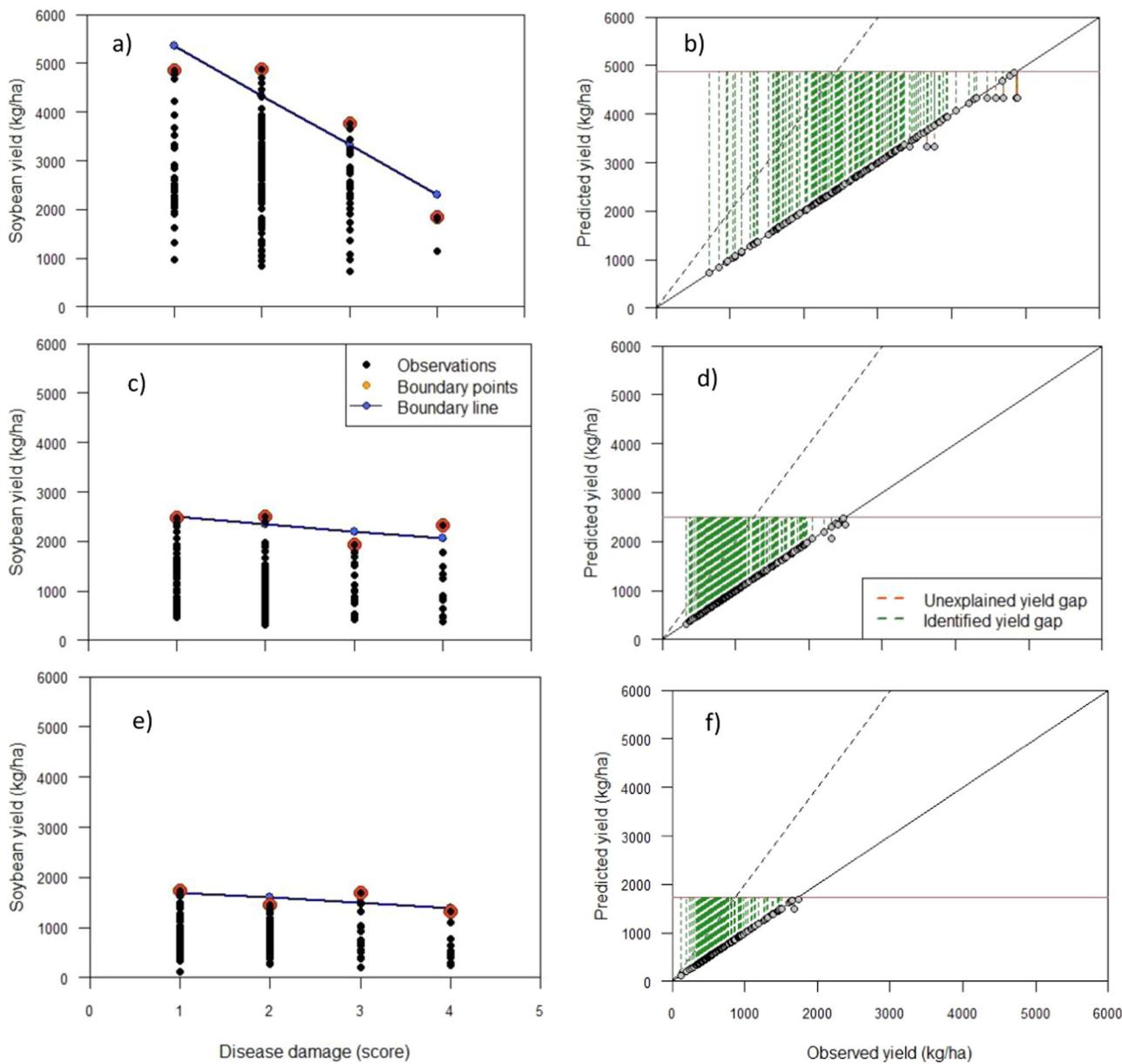


Fig. 4 | Disease damage effect on soybean yield with a fitted boundary line, and yield gap attributed to disease damage. **a** The fitted boundary line, and **b** the yield gap associated with disease damage in Malawi; likewise for **c, d** in Zambia; equally for **e, f** in Mozambique. The scores for disease damage: 1—no sign of disease damage,

2—limited signs of disease damage, 3—moderate signs of disease damage, 4—widespread sign of disease damage. The 1:1 represents when the actual yield is equal to the predicted yield. The points that fall below the 1:1 line indicate yield gap that cannot be explained by disease damage.

ridges spaced at 75 cm apart and on each they make two grooves (20–30 cm apart) to plant soybeans⁵³. This minimizes competition among the plants. Additionally, the seed rate recommendation in Malawi and Zambia by the government is 60–65 kg ha⁻¹⁵³ and 80–110 kg ha⁻¹ depending on the variety⁵⁴, respectively compared to 50–60 kg ha⁻¹ in Mozambique. Moreover, the seed viability in Mozambique could be low due to the recycling of previously certified seed, a question that wasn't clearly elaborated in the survey questionnaire. Malita et al.⁵⁵ confirmed that most farmers regularly recycle legume seeds which leads to a decline in viability⁵⁶ causing repeated gapping and hence higher seed rate. This probably occurred in Mozambique due to poor soil nutrients or moisture⁵⁷, a scenario corroborated by early planting with minimum soil moisture and low application of fertilizers⁵⁸, especially phosphorus to soybean⁵⁹—also supported by the survey results of this study in which only a few farmers apply fertilizer to soybeans.

Across the Southern Africa region (Malawi, Zambia, and Mozambique), besides soil fertility and fertilizer application which was not determined

in this survey, soybean variety causes a higher yield gap, followed by the seed rate, and then disease damage, especially soybean rust⁴. Despite breeding efforts and the release of promiscuous varieties i.e., the Tropical glycine cross (TGx) lines from the International Institute of Tropical Agriculture²⁹ and other seed companies, the soybean seed system is still plagued by many problems. Mainly the price of certified improved seeds is high so most farmers can't afford them, lack of clear distribution channels and knowledge of suitable site-specific varieties⁶⁰. Interestingly from the results of this study, some countries have higher seed rates than the government recommendation. An indication of availability of seed, but most likely recycled seed that has low viability which requires a high seed rate to obtain the desired plant population. Indeed recycled seeds have reduced viability^{56,61}.

Minimizing yield gap is key to increasing yield per unit area and agronomic gain, however, it is necessary to understand the yield-limiting factors and prioritize interventions considering the few resources that smallholder farmers have. This study has shown that to close the soybean

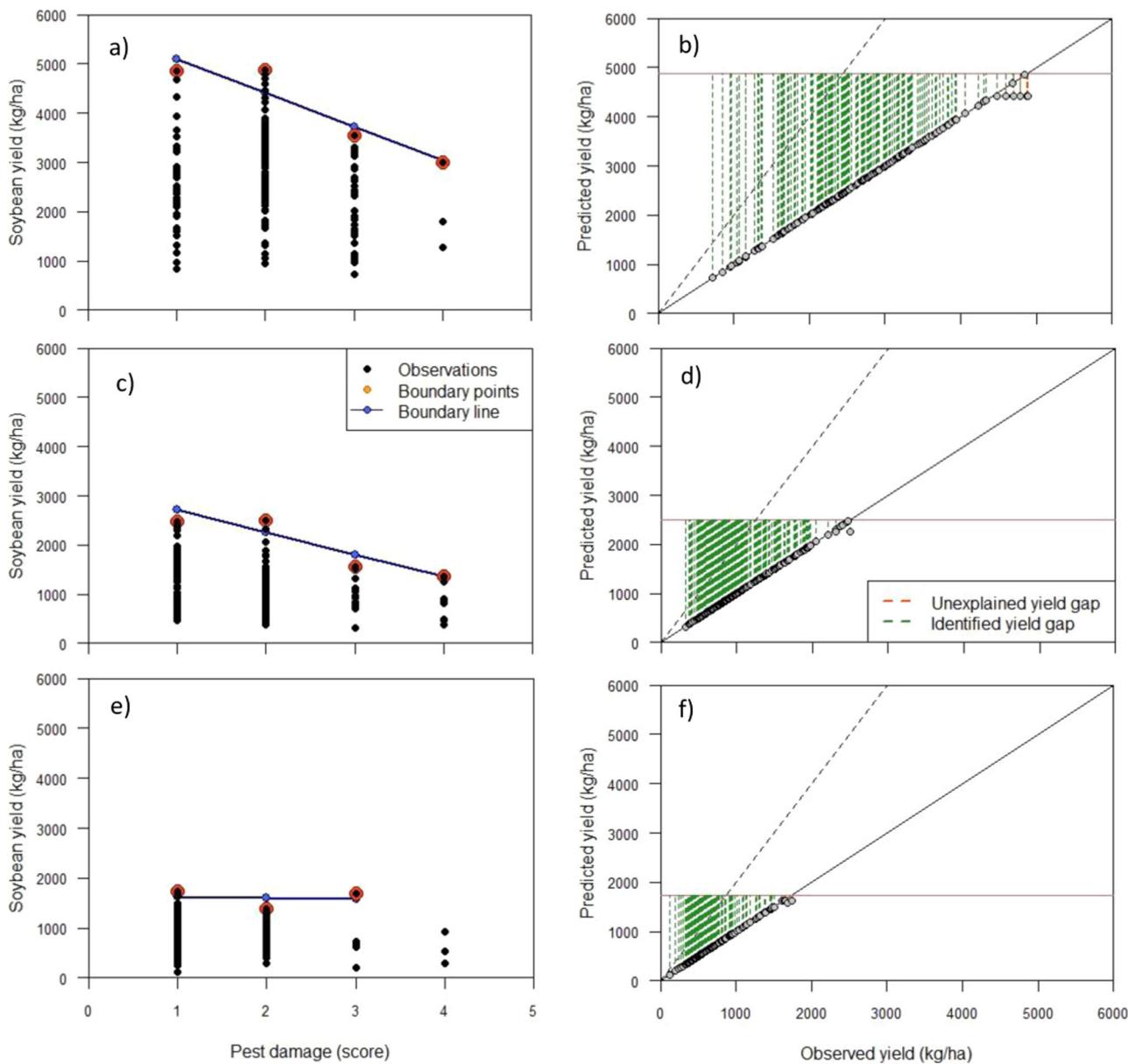


Fig. 5 | The impact of pest damage on soybean yield with a fitted boundary line, and the yield gap related to pest damage. **a** The fitted boundary line and **b** the yield gap associated with pest damage of soybean in Malawi; similarly for **c, d**, respectively, in Zambia; also for **e, f**, respectively, in Mozambique. The scores for pest damage: 1—

no sign of pest damage, 2—limited signs of pest damage, 3—moderate signs of pest damage, 4—widespread sign of pest damage. The 1:1 represents when the actual yield is equal to the predicted yield. The points that fall below the 1:1 line indicate yield gap that cannot be explained by pest damage.

yield gap, besides soil nutrients, soybean variety is the most limiting factor in Malawi and Zambia followed by disease damage in Malawi and seed rate in Zambia. Whereas, in Mozambique, seed rate is significant, although the seeds should have high viability, followed by the variety. Overall, in the Southern Africa region (Malawi, Zambia, and Mozambique) the major soybean yield gap contributors are: variety, seed rate, and disease damage, especially soybean rust, in that order.

Release of improved high-yielding promiscuous soybean varieties like Chitedze 4 and Tikolore in Malawi, Wamini and Kafue in Mozambique and Zambia, respectively, notwithstanding, most farmers are yet to adopt suitable varieties for their specific regions. And even those who have adopted them tend to recycle the seeds leading low growth vigour and poor yields. Thus, there is need to improve certified seeds production through the community-based seed producers, enhance its access and availability and quality monitoring – essentially strengthening the soybean seed production system with a focus on the end-user benefits. Secondly, and linked to the

seed systems, seed rate is mostly lower due to recycling and poor germination, which can be addressed by strengthening the seed system, ensuring appropriate planting time adhering to effective soil moisture, and suitable planting population. Thirdly, disease surveillance needs strengthening, with the spread and help of smartphones and feature phones, farmers self-reporting of disease outbreaks like soybean rust could be tracked via social media, and real-time solutions offered to prevent further yield losses.

Methods

Sites

The data was collected in two districts of Malawi (Lilongwe and Kasungu) from farmers located between 12.82° – 13.97° S and 33.44° – 33.74° E; in Zambia in Chipata and Sinda districts (14.21° – 14.20° S and 31.85° – 32.22° E), and in Mozambique in Angonia district (14.67° – 14.73° S and 33.96° – 34.38° E) in 2023. These areas are at an altitude of 915–1557 m above sea level, with the region predominantly referred to as the Chinyanja Triangle

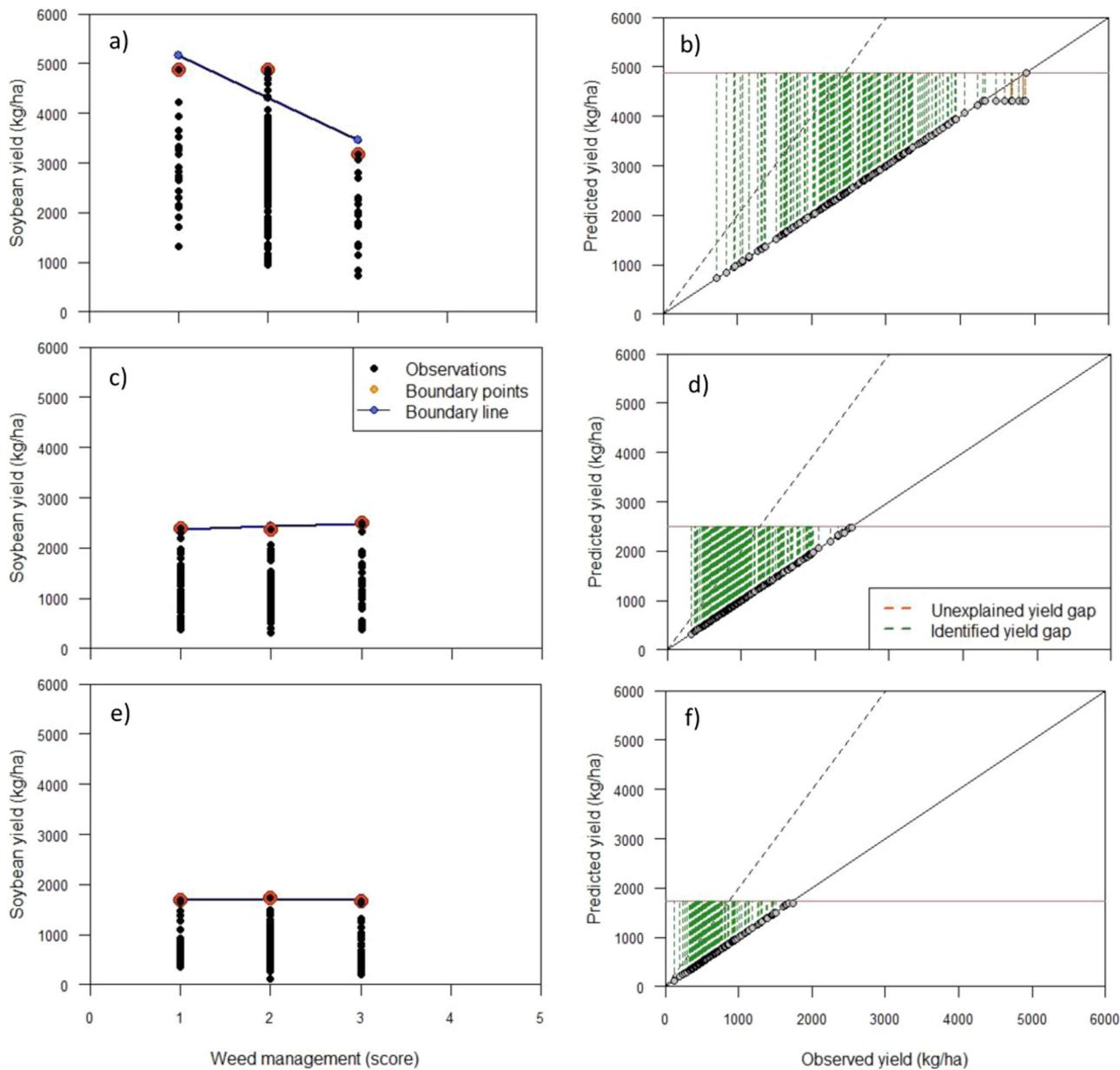


Fig. 6 | Weeding influence on soybean yield with a fitted boundary line, and the yield gap attributed to weed management option. a, c, e The fitted boundary line, and **b, d, f** the yield gap arising due to weed management in Malawi, Zambia, and Mozambique, respectively. The scores for weed management: 1—well managed (no

weed infestation), 2—moderately managed (some weeds observed), 3—poorly managed (>half the field is infested with weeds). The 1:1 represents when the actual yield is equal to the predicted yield. The points that fall below the 1:1 line indicate yield gap that cannot be explained by weed infestation.

(Fig. 7a). In this region, the average and maximum atmospheric temperature, rainfall amount, and wind speed from 1st October 2022 to 30th April 2023, the period during which most of the farmers reported soybean being in the field (from planting to harvesting) is presented in Fig. 7b–e. This area is within the Zambezi River Basin and lies between the Luangwa River on the West, Lake Malawi, and River Shire on the East, and Zambezi river on the South. It is popular with the Chichewa/Chinyanja-speaking people sharing a similar history, language, and culture across the Eastern province of Zambia, the Central regions of Malawi, and the Tete Province of Mozambique. The three communities in the three countries are differentiated by biophysical conditions, as well as country-specific political and economic conditions. The Chinyanja Triangle is dominated by maize-based farming systems.

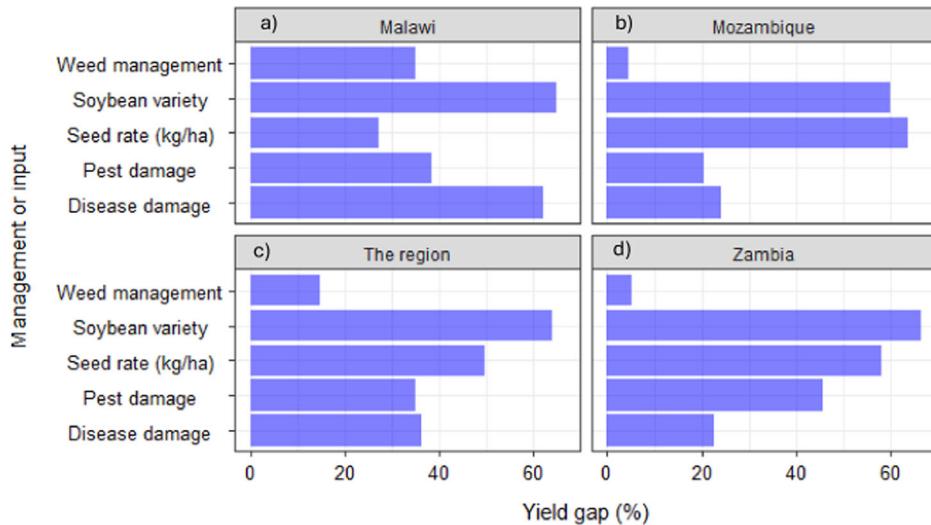
Data collection and crop cut survey

Crop cut surveys generate production data through direct measurement and it is widely used as a method of yield estimation. The major

objective of a crop-cut survey is to measure actual crop yields under actual farmer management conditions through randomly selected sample plots within a crop field. From the selected plot, a sub-plot is harvested and weighed appropriately. Then, the yield per unit area is calculated as the weight of the harvested crop yield adjusted for grain moisture content⁶².

The yield cut survey was conducted in Malawi, Zambia, and Mozambique following the protocol by CIMMYT (2023). In each country, 180 farmers/households growing soybeans (Fig. 7) responded to the survey questions which were administered by trained enumerators in a standardized questionnaire through ODK⁶³ and uploaded on ONA⁶⁴. In Malawi, data was collected from 120 households in Kasungu and 60 households in Lilongwe districts, whereas in Zambia 120 households were from Katete and 60 from Sinda districts. In Mozambique, all the 180 households were from Angonia district. The survey questions were on administrative levels, farmer name and gender, location, land size, crop grown and management

Fig. 7 | Yield gap due to various management or input options. In **a** Malawi, **b** Mozambique, **c** Zambia, and **d** in all the three countries (Malawi, Zambia, and Mozambique).



practices (weeding, pest and disease control), actual yield cut for yield calculation, inputs, inputs and produce prices.

Disease damage was evaluated in the three countries by scoring: no sign of disease damage as '1', '2' as limited signs of disease damage, '3' as moderate signs of disease damage, and '4' as a widespread sign of disease damage. In the three countries, pest damage was scored using a scale of '1' as no sign of pest damage, '2' as limited signs of pest damage, '3' as moderate signs of pest damage, and '4' represented widespread signs of pest damage. Weed management impact in the three countries was scored on a scale of '1' representing well managed (no weed infestation), '2' being moderately managed (some weeds observed), and '3' as poorly managed (>half the field is infested with weeds).

In the yield cut protocol⁶², GPS coordinates, and area measurements were embedded in the ODK tool as part of the survey. The farmers were registered with unique farm identifiers (IDs) for each field and their plot areas were measured using the polygon option by walking around the perimeter of the plot. The field was then divided into four quadrants assuming that it was a rectangle. This was followed by assuming a diagonal line from the northwest (NW) to the southeast corners (SE) placing three sub-plots of 2 m by 2 m where the actual crop cut was to be conducted, the first on the top left quadrant, the second at the center, and the last on the bottom right. After this, the above-ground part of soybean was cut within each 2 m by 2 m sub-plot, the pods were separated from the haulms, threshed and the grains weighed. The fresh weight of the grains was recorded and later dry weight to 12% moisture content. The dry weights were used to calculate the yield per hectare for each field/farmer by averaging the three 2 m by 2 m sub-plots.

Yield gap decomposition using boundary line analysis

In the boundary line analysis (BLA), the water-limited yield (Y_w) is derived from the highest farmers' yield which is defined as the average top 10% of farmers' yields whereas the remaining 90% forms the actual farmers' yields (Y_a) that can be achieved for a given input level in a well-defined biophysical environment.

Various parameters were collected during the survey, however, those subjected to boundary line analysis were yield (kg ha^{-1}), seed rate (kg ha^{-1}), soybean variety, disease and pest damage, and weed management.

The continuous variables such as seed rate were analysed, whereas the categorical variables/management variables such as weed management, disease and pest damage, and variety type were first given numerical codes before analysis.

Even though there are no generally agreed protocols for BLA, Hajjarpoor et al.⁶⁵ outlined the following important processes:

1. Examining the scatter plot of data: a scatter plot (XY chart) should be prepared with crop yield as the dependent variable and one selected independent variable (e.g., weeding).
2. Removal of the outliers.
3. Selection of the data points from the upper limit of the data for curve fitting.
4. Fitting a function to the data points in 2.

The above processes were conducted in R⁶⁶. The R packages installed and loaded included for the analysis included "splines", "metrics", "dplyr", "tidy", "knitr", "reshape2", "ggplot2", "mass", "ggstatsplot". After data was uploaded and boxplot analysis was conducted to identify outliers and eliminate them, boundary line models are susceptible to the effects of outliers⁶⁵. The boundary line (BL) point function was defined, and the 'X' variable of interest—selected dry grain yield of soybean in our case. Using the binning approach, which is a heuristic approach, the 'X' variables were split into ten quantiles (from 0.1 to 1.0), and BL points for each quantile were defined based on maximum value or on the yields when approaching the 95th quantile. The highest farmers' yields were set at the average of top 10% of the actual farmers' yields to represent the Y_w while the rest were Y_a for every site i.e., each site was treated individually assuming that its biophysical characteristics are unique to itself, for example, the values for Kasungu in Malawi were calculated individually and not combined with those from either Angonia in Mozambique or Katete in Zambia. This was to ensure that yields within a given agro-ecology are treated the same and no introduction of variability due to biophysical differences from other regions. After setting the boundary points, the 'Y' variables were defined, and yield decomposition analysis was performed for each variable. The boundary line was fitted: for the continuous variables like seed rate and the discrete variables like disease damage in this study, Gaussian General Linear Model and Poisson General Linear Model were used, respectively. The results were then plotted as yield gap indicating the boundary points and the boundary line. Followed by the determination and the plotting of the identified yield gap (IYG), and unexplained yield gap (UYG). The yield gap estimation and decomposition were conducted by plotting the predicted yield (estimated using the boundary lines) against the actual yields. The continuous red line on each observed versus predicted yield graph represents the highest yield observed in the study area and it is taken as the attainable yield. The 1:1 line represents the situation when the actual yield is equal to the predicted yield. The points that fall below the 1:1 line indicate the unexplained yield gap i.e., the yield difference between the predicted and attainable yield that can not be explained by the most limiting factor.

After fitting the boundary lines, the percent yield gap for each site was calculated as follows:

$$\text{Yield gap (\%)} = \left(\frac{Y_w - Y_a}{Y_w} \right) \times 100\% \quad (1)$$

Data availability

Datasets are available from the corresponding author on reasonable request.

Code availability

Codes can be provided on request from the corresponding author.

Received: 5 June 2024; Accepted: 2 December 2024;

Published online: 31 December 2024

References

- Agomoh, I. V., Drury, C. F., Yang, X., Phillips, L. A. & Reynolds, W. D. Crop rotation enhances soybean yields and soil health indicators. *Soil Sci. Soc. Am. J.* **85**, 1185–1195 (2021).
- Ali, W., Ahmad, M. M., Iftikhar, F., Qureshi, M. & Ceyhan, A. Nutritive potentials of soybean and its significance for humans health and animal production: a review. *Eurasian J. Food Sci. Technol.* **4**, 41–53 (2020).
- Tufa, A. H. et al. The productivity and income effects of adoption of improved soybean varieties and agronomic practices in Malawi. *World Dev.* **124**, 104631 (2019).
- Murithi, H. M., Beed, F., Tukamuhabwa, P., Thomma, B. P. H. J. & Joosten, M. H. A. J. Soybean production in eastern and southern Africa and threat of yield loss due to soybean rust caused by *Phakopsora pachyrhizi*. *Plant Pathol.* **65**, 176–188 (2016).
- Omondi, J. O. et al. Closing the yield gap of soybean (*Glycine max* (L.) Merrill) in Southern Africa: a case of Malawi, Zambia, and Mozambique. *Front. Agron.* **5**, 1–14 (2023).
- AGRA. Africa Agriculture Status Report: The Business of Smallholder Agriculture in Sub-Saharan Africa. Alliance for a Green Revolution in Africa <https://agra.org/aasr2017/preface/> (2017).
- Van Vugt, D., Franke, A. C. & Giller, K. E. Participatory research to close the soybean yield gap on smallholder farms in Malawi. *Exp. Agric.* **53**, 396–415 (2017).
- Grant, W., Wolfaardt, A. & Louw, A. Maize Value Chain in the SADC Region. A technical report by AECOM International Development. USAID Contract No. 674-C-00-10-00075-00 (2012).
- Jakobsen, J. & Westengen, O. T. The imperial maize assemblage: maize dialectics in Malawi and India. *J. Peasant Stud.* **49**, 536–560 (2022).
- Khojely, D. M., Ibrahim, S. E., Sapey, E. & Han, T. History, current status, and prospects of soybean production and research in sub-Saharan Africa. *Crop J.* **6**, 226–235 (2018).
- Foyer, C. H. et al. Modelling predicts that soybean is poised to dominate crop production across Africa. *Plant, Cell Environ.* **42**, 373–385 (2019).
- Nsomba, G., Roberts, S., Tshabalala, N. & Manjengwa, E. Assessing Agriculture & Food Markets in Eastern and Southern Africa: An Agenda for Regional Competition Enforcement. CCRED African Market Observatory Working Paper (2022).
- Ncube, P., Roberts, S., Zengeni, T. & Samboko, P. C. Identifying Growth Opportunities in the Southern African Development Community through Regional Value Chains: The Case of the Animal Feed to Poultry Value Chain. CCRED Working Paper No. 8/2016, WIDER Working Paper 2017/4. (2017).
- Mhango, W. G., Snapp, S. S. & Phiri, G. Y. K. Opportunities and constraints to legume diversification for sustainable maize production on smallholder farms in Malawi. *Renew. Agric. Food Syst.* **28**, 234–244 (2012).
- Renwick, L. L. R. et al. Long-term crop rotation diversification enhances maize drought resistance through soil organic matter. *Environ. Res.* **16**, 084067 (2021).
- Siamabele, B. The significance of soybean production in the face of changing climates in Africa. *Cogent Food Agric.* **7**, 1933745 (2021).
- Kolapo, A. L. Soybean: Africa's potential Cinderella food crop. in *Soybean - biochemistry, chemistry and physiology* (ed. Ng, T.-B.) 642 (Intech, 2011).
- Ciampitti, I. A. et al. Revisiting biological nitrogen fixation dynamics in soybeans. *Front. Plant Sci.* **12**, 727021 (2021).
- Matova, P. M., Kamutando, CasperN., NKutywayo, D., Gutsa, F. & Labuschagne, M. T. Fall-armyworm invasion, control practices and resistance breeding in Sub-Saharan Africa. *Crop Sci.* **60**, 2951–2970 (2020).
- Hailu, G., Niassy, S., Zeyaur, K. R., Ochatum, N. & Subramanian, S. Maize-legume intercropping and push-pull for management of fall armyworm, stemborers, and striga in Uganda. *Agron. J.* **110**, 2513–2522 (2018).
- Muyanga, M. & Jayne, T. S. Effects of rising rural population density on smallholder agriculture in Kenya. *Food Policy* **48**, 98–113 (2014).
- Headey, D. D. & Jayne, T. S. Adaptation to land constraints: Is Africa different? *Food Policy* **48**, 18–33 (2014).
- Mellor, J. W. High rural population density Africa – What are the growth requirements and who participates? *Food Policy* **48**, 66–75 (2014).
- Van Ittersum, M. K. et al. Yield gap analysis with local to global relevance-A review. *F. Crop. Res.* **143**, 4–17 (2013).
- Silva, J. V. et al. Narrowing maize yield gaps across smallholder farming systems in Zambia: what interventions, where, and for whom? *Agron. Sustain. Dev.* **43**, 26 (2023).
- Minussi Winck, J. E. et al. Decomposition of yield gap of soybean in environment x genetics x management in Southern Brazil. *Eur. J. Agron.* **145**, 126795 (2023).
- Mbanya, W. Assessment of the constraints in soybean production: a case of Northern Region, Ghana. *J. Dev. Sustain. Agric.* **6**, 199–214 (2011).
- Miti, C., Milne, A. E., Giller, K. E. & Lark, R. M. The concepts and quantification of yield gap using boundary lines. A review. *F. Crop. Res.* **311**, 109365 (2024).
- Chigeza, G. et al. Public sector soybean (*Glycine max*) breeding: advances in cultivar development in the African tropics. *Plant Breed.* **138**, 455–464 (2019).
- Abate, T. *Four Seasons of Learning and Engaging Smallholder Farmers: Progress of Phase 1.* (Springer Nature, 2012).
- Rittler, L. & Bykova, O. Water use and irrigation in soybean. *Legum. Transl.* 221–255 (2022) <https://doi.org/10.1002/9780470751114.ch10>.
- Adjei-Nsiah, S., Martei, D., Yakubu, A. & Ulzen, J. Soybean (*Glycine max* L. Merrill) responds to phosphorus application and rhizobium inoculation on Acrisols of the semi-deciduous forest agro-ecological zone of Ghana. *PeerJ* **10**, e12671 (2022).
- He, J. et al. Phosphorus application increases root growth, improves daily water use during the reproductive stage, and increases grain yield in soybean subjected to water shortage. *Environ. Exp. Bot.* **166**, 103816 (2019).
- Briat, J. F., Gojon, A., Plassard, C., Rouached, H. & Lemaire, G. Reappraisal of the central role of soil nutrient availability in nutrient management in light of recent advances in plant nutrition at crop and molecular levels. *Eur. J. Agron.* **116**, 126069 (2020).
- Viano, A. Gene editing solution to reducing losses due to soybean rust. 12 <https://zodiakmalawi.com/70-special-reports/6471-gene-editing-solution-to-reducing-losses-due-to-soybean-rust#:~:text=Despite> its huge potential%2C the, majority farmers in the cold. (2023).
- Li, X. & Yang, X. B. Similarity, pattern, and grouping of soybean fungal diseases in the United States: implications for the risk of soybean rust. *Plant Dis.* **93**, 162–169 (2009).

37. Levy, C. Epidemiology and chemical control of soybean rust in southern Africa. *Plant Dis.* **89**, 669–674 (2005).

38. Diao, X., Silver, J. & Takeshima, H. Agricultural Mechanization and Agricultural Transformation. African Transformation Report 2016: Transforming Africa's Agriculture. Joint research between: African Center for Economic Transformation (ACET) and Japan International Cooperation Agency Research institute (JICA-RI). (2016).

39. Meyer, F. et al. *Modelling Soybean Markets in Eastern and Southern Africa*. <https://ec.europa.eu/jrc/https://doi.org/10.2760/20598> (2018).

40. Neto, V. A. L. et al. Fungicide management in the Asian soybean rust. *Rev. Agrogeoambiental* **13**, 36–46 (2021).

41. Isard, S. A., Russo, J. M. & Ariatti, A. The integrated aerobiology modeling system applied to the spread of soybean rust into the Ohio River valley during September 2006. *Aerobiologia* **23**, 271–282 (2007).

42. Li, X., Yang, X. & Mo, J. Estimation of soybean rust uredospore terminal velocity, dry deposition, and the wet deposition associated with rainfall. *Eur. J. Plant Pathol.* **123**, 377–386 (2009).

43. Hossain, M. M. et al. Understanding Phakopsora pachyrhizi in soybean: comprehensive insights, threats, and interventions from the Asian perspective. *Front. Microbiol.* **14**, 1–20 (2023).

44. Reis, E. M., Guerra, W. D., Zanatta, M. & Zambolim, L. An overview of the epidemiology of Phakopsora pachyrhizi in the State of Mato Grosso, Brazil. *J. Agric. Sci.* **13**, 110 (2021).

45. Harre, N. T. & Young, B. G. Early-season nutrient competition between weeds and soybean. *J. Plant Nutr.* **43**, 1887–1906 (2020).

46. Michaud, J. P. & Grant, A. K. The biology and behavior of the longhorned beetle, *Dectes texanus* on sunflower and soybean. *J. Insect Sci.* **5**, 1–15 (2005).

47. Andrade, K. et al. Bioecological characteristics of *Chrysodeixis includens* (Lepidoptera: Noctuidae) fed on different hosts. *Austral Entomol.* **55**, 449–454 (2016).

48. Pant, C. & Sah, S. K. Managing plant population and competition in field crops. *Acta Sci. Malaysia* **4**, 57–60 (2020).

49. Craine, J. M. & Dybzinski, R. Mechanisms of plant competition for nutrients, water and light. *Funct. Ecol.* **27**, 833–840 (2013).

50. Haarhoff, S. J. & Swanepoel, P. A. Plant population and row spacing affects growth and yield of rainfed maize in semi-arid environments. *Front. Plant Sci.* **13**, 761121 (2022).

51. Silva, M. et al. Increasing population density reduces soybean yield components and productivity. *Biosci. J.* **37**, e37042 (2021).

52. Savala, C. E. N., Wiredu, A. N., Okoth, J. O. & Kyei-Boahen, S. Inoculant, nitrogen and phosphorus improves photosynthesis and water-use efficiency in soybean production. *J. Agric. Sci.* **159**, 349–362 (2021).

53. Kananji, G. A. D. et al. *A Guide to Soybean Production in Malawi*. <https://doi.org/10.13140/2.1.4982.2723> (2013).

54. Ministry of Agriculture. Enhanced Smallholder Agribusiness Promotion Programme: Soybeans Production Manual. <https://www.ifad.org/en/web/operations/project/id/2000001405/country/zambia> (2019).

55. Malita, C. et al. *Feed the Future Mozambique Improved Seeds for Better Agriculture (SEMEAR)*. https://pdf.usaid.gov/pdf_docs/PA00WDS4.pdf (2019).

56. Nagel, M. & Börner, A. The longevity of crop seeds stored under ambient conditions. *Seed Sci. Res.* **20**, 1–12 (2010).

57. Lamichhane, J. R. et al. Analysis of soybean germination, emergence, and prediction of a possible northward establishment of the crop under climate change. *Eur. J. Agron.* **113**, 125972 (2020).

58. Bationo, A. et al. *Fighting Poverty in Sub-Saharan Africa: The Multiple Roles of Legumes in Integrated Soil Fertility Management*. (Springer Dordrecht, 2011).

59. Vanlauwe, B., Hungria, M., Kanampiu, F. & Giller, K. E. The role of legumes in the sustainable intensification of African smallholder agriculture: lessons learnt and challenges for the future. *Agric. Ecosyst. Environ.* **284**, 106583 (2019).

60. Farrow, A. et al. From best fit technologies to best fit scaling: Incorporating and evaluating factors affecting the adoption of grain legumes in Sub-Saharan Africa. *Exp. Agric.* **55**, 226–251 (2019).

61. Cortelazzo, A. L., Coutinho, J. & Granjeiro, P. A. Storage and ageing of french beans (*Phaseolus vulgaris* L.): effect on seed viability and vigor. *Braz. J. Morphol. Sci.* **22**, 121–128 (2005).

62. Kindie, T., Tesfaye, S., Lulseged, T., Gizaw, D. & Wuletawu, A. Crop Cut Surveys Yield Gap Decomposition Add-on. *Excell. Agron. Initiat. CGIAR*. (2023).

63. ODK. *Collect data anywhere*. at <https://getodk.org/> (2020).

64. ONA. Where need is greatest, using data to transform. at <https://ona.io/home/> (2024).

65. Hajarpoor, A. et al. Using boundary line analysis to assess the on-farm crop yield gap of wheat. *F. Crop. Res.* **225**, 64–73 (2018).

66. R Core Team. A language and environment for statistical computing. *R Foundation for Statistical Computing* at <https://www.r-project.org/> (2023).

Acknowledgements

The authors are grateful to One CGIAR for funding this work through the Excellence in Agronomy (EiA) and Sustainable Intensification - Mixed Farming Initiative (SI-MFS). Their gratitude also goes to colleagues at Solidaridad, IITA, CIMMYT, and ICRAF.

Author contributions

J.O.O., M.S.C., P.M., and I.N. conceived and designed the manuscript. J.O.O. and M.S.C. collected the data. J.O.O. synthesized, analysed, and presented the data. J.O.O. drafted the manuscript, it was reviewed by M.S.C., P.M., and I.N. and S.K.B. J.O.O. edited and revised the manuscript's final version.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to John Okoth Omondi.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024, corrected publication 2025