Rothamsted Research
Harpenden, Herts, AL5 2JQ

Telephone: +44 (0)1582 763133
Web: http://www.rothamsted.ac.uk/

MacLaren, C., Mead, A., Van Balen, D., Claessens, L., Etana, A., De Haan, J., Haagsma, W., Jack, O., Keller, T., Labuschagne, J., Myrbeck, A., Necpalova, M., Nziguheba, G., Six, J., Strauss, J., Swanepoel, P. A., Thierfelder, C., Topp, C., Tshuma, F., Verstegen, H., Walker, R., Watson, C., Wesselink, M. and Storkey, J. 2022. Long-term evidence for ecological intensification as a pathway to sustainable agriculture. *Nature Sustainability.* https://doi.org/10.1038/s41893-022-00911-x

Long-term evidence for ecological intensification as a pathway to sustainable agriculture

Chloe MacLaren*[1,2], Andrew Mead[3], Derk van Balen[4], Lieven Claessens[4,5], Ararso Etana[6], Janjo de Haan[4], Wiepie Haagsma[4], Ortrud Jäck[7], Thomas Keller[6,8], Johan Labuschagne[9], Åsa Myrbeck[6,10], Magdalena Necpalova[11,12], Generose Nziguheba[13], Johan Six[12], Johann Strauss[2,9], Pieter Andreas Swanepoel[2], Christian Thierfelder[14], Cairistiona Topp[15], Flackson Tshuma[2], Harry Verstegen[4], Robin Walker[16], Christine Watson[7,16], Marie Wesselink[4], Jonathan Storkey[1]

** chloe.maclaren@rothamsted.ac.uk*

[1]Protecting Crops and the Environment, Rothamsted Research, West Common, Harpenden AL5 2JQ, UK
[2]Department of Agronomy, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa
[3]Intelligent Data Ecosystems, Rothamsted Research, West Common, Harpenden AL5 2JQ, UK
[4]Field Crops, Wageningen University and Research, P.O. Box 430, 8200 AK Lelystad, The Netherlands[5]International Institute of Tropical Agriculture (IITA), P.O. Box 10, Duluti, Arusha, Tanzania
[6]Department of Soil and Environment, Swedish University of Agricultural Sciences (SLU), Box 7014, 75007 Uppsala, Sweden
[7]Department of Crop Production Ecology, Swedish University of Agricultural Sciences (SLU), Box 7043, 75007 Uppsala, Sweden
[8]Department of Agroecology and Environment, Agroscope, Reckenholzstrasse 191, 8046 Zürich, Switzerland
[9]Western Cape Department of Agriculture, Private Bag X1, Elsenburg, 7607, South Africa
[10] Department of Agriculture and Food, Research Institutes of Sweden (RISE), Box 857, 501 15 Borås, Sweden
[11]Environment & Sustainable Resource Management Section, School of Agriculture & Food Science, University College Dublin, Ireland
[12]Department of Environmental Systems Science, ETH Zurich, Universitätstrasse 2, 8092 Zurich, Switzerland
[13]International Institute of Tropical Agriculture (IITA), c/o ICIPE, Kasarani, Off Thika Highway, PO 30772, Nairobi 00100, Kenya
[14]Southern Africa Regional Office, International Maize and Wheat Improvement Centre (CIMMYT), P.O. Box MP163, Harare, Zimbabwe
[15]SRUC Edinburgh, West Mains Road, Edinburgh EH9 3JG, UK
[16]SRUC Aberdeen, Craibstone Estate, Aberdeen AB21 9YA, UK

Abstract

Ecological intensification could help return agriculture into a ‘safe operating space’ for humanity. Using a novel application of meta-analysis to data from 30 long-term experiments from Europe and Africa (comprising 25,565 yield records), we investigated how field-scale EI practices interact with each other, and with N fertiliser and tillage, in their effects on long-term crop yields. Here we confirmed that EI practices (specifically increasing crop diversity, adding fertility crops, and organic matter) have generally positive effects on the yield of staple crops. However, we show that EI practices have a largely substitutive interaction with N fertiliser, so that EI practices substantially increase yield at low N fertiliser doses, but have minimal or no effect on yield at high N fertiliser doses. EI practices had comparable effects across different tillage intensities and reducing tillage did not strongly affect yields.

Main

Agriculture is a leading cause of global environmental change, while also being highly vulnerable to that change[1]. Human activities, including agriculture, have increased greenhouse gas emissions, nutrient bioavailability, habitat loss, and species extinctions toward ‘planetary boundaries’, where Earth’s environment is at high risk of shifting to a less hospitable state[2,3]. This in turn threatens agriculture through increasing the likelihood of extreme weather events, resource depletion, and pest outbreaks[4,5]. Agriculture must address these environmental challenges whilst also meeting the needs of a growing global population. Although many political and societal changes could limit future food demand (such as fairer food distribution and reduced animal product consumption[6,7]), it must also be assumed that yields of the world’s staple crops will, at the very least, need to be maintained[8].

Ecological Intensification (EI) is one pathway proposed to sustain yields while reducing adverse impacts of agriculture on the environment (and consequently reducing threats posed to agriculture *by* the environment). EI is defined as the enhancement of ecosystem services[9] to complement or substitute for the role of anthropogenic inputs in maintaining or increasing yields[10,11]. Anthropogenic inputs have underpinned necessary gains in productivity and food security since the Green Revolution, but their widespread over-use has incurred substantial environmental costs[12]. EI seeks to retain productivity whilst mitigating environmental impacts, and is a strategy that could be implemented under various sustainable agriculture paradigms, such as agroecology[13], sustainable intensification[14], and Climate Smart Agriculture[15]. Managing farmland to provide ecosystem services that support productivity can also encourage farmers to avoid environmentally degrading practices – leading Tittonell[16] to describe EI as both “sustained by nature and sustainable in nature”.

Here, we investigate the extent to which crop yields can be supported by field-scale EI practices targeted at enhancing the ecosystem services of nutrient cycling and regulating weeds, pests, and diseases. Input-based, field-scale practices to achieve high yields involve regular and intensive inputs of tillage, synthetic fertilisers, and pesticides, which together can lead to increased carbon emissions and the release of pollutants and soil particulates into surrounding habitats[17,18]. Identifying and upscaling farming practices that decouple high yields from high use of these inputs would therefore facilitate returning to a global ‘safe operating space’[2,7]. There is promising evidence that many field-scale EI practices could contribute to this decoupling[11], such as using legumes to fix nitrogen[19], diversifying crops to better regulate weeds, pests and diseases[20], recycling manures to fertilise crops[21], and managing crop residues to improve soil quality[22].

Realising the full potential for EI, however, requires knowledge of the relative yield response to different EI practices and inputs, and the extent to which this response is context-dependent. The aim of EI may differ depending on the context; for example, in a high-input high-yield scenario, EI practices may be intended to reduce inputs and thus environmental impacts while sustaining yields, to bring cropping systems back within a global 'safe operating space'[2]. In low-yield low-input systems, EI practices might improve food security by complementing inputs to increase yields, in the face of low input accessibility[31] or adverse local conditions[32]. However, it is important to understand whether different EI practices have different effects in these different contexts, so that the optimal combinations of EI practices and inputs can be used to achieve the desired aim.

The over-arching picture of the relative effects of and interactions between EI practices and inputs has so far remained unclear, because it is challenging for individual experiments to test more than one or two practices or inputs in concert (given the need for enough area to replicate multiple treatments). Meta-analyses can compare relative effects across multiple experiments[23] but have not yet been applied to explore whether different EI practices and inputs interact in their effects on yield. Previous research in EI has also been limited by the short term focus of many studies that address effects on a single crop over one or two years[11], while the true impacts of different agronomic practices may only become apparent over long time scales when the effects of interannual variability, short-term perturbations, and transitional dynamics can be accounted for[24,25].

To address this knowledge gap, this study collated data across 30 long-term experiments (LTEs) in Europe and Africa (with a minimum age of 9 years) to investigate the relative yield effects of different EI practices and inputs. Analyses of multiple LTEs have previously been used to quantify the effect of crop diversification on yields[26,27] and to compare different soil management practices[23], but not yet to explore interactions among multiple EI practices and inputs. Together, the LTEs in our dataset assess three different EI practices: (1) crop diversification from a monoculture (CD), (2) addition of 'fertility' crops to an arable rotation (FC) and (3) organic matter management (OM; including soil amendments and crop residues) (Table 1). Each of these offers opportunities to increase ecological functioning by increasing diversity and/or connecting resource flows within and/or between farmed fields[19–21]. Many LTEs tested EI practices alongside different synthetic nitrogen fertiliser application rates (NF) and tillage intensities (TI), allowing us to investigate the effects of EI practices (and combinations thereof) at different levels of these inputs. We consider tillage intensity to be an anthropogenic input of energy and disturbance, which incurs fuel use and soil degradation[18,28]. Other inputs of high potential interest in relation to field-scale EI practices are phosphorus[29] and pesticides[30], but we had insufficient data to assess these.

In total, our dataset consisted of 25,565 plot-by-year yield records. While individual results have been published for most LTEs, here we further realise the potential of these LTEs by synthesising data across experiments to test overarching hypotheses. To combine evidence for multiple practices across multiple LTEs with contrasting cropping systems and treatment structures, it was necessary to develop a novel meta-analysis procedure to directly quantify the association between the relative yields and differences in each EI practice and input. Our specific objectives for this analysis were to i) quantify the relative yield response to different EI practices and inputs in different combinations, ii) use these results to assess the potential for EI practices to increase crop yields for a given level of inputs, or to sustain yields at reduced levels of inputs.

Exploring EI via meta-analysis of multiple LTEs

To explore the relative effects of different EI practices on yield across the 30 LTEs (Supplementary Table S1.1), we used a novel three-step procedure to integrate data from experiments with different crops and different treatment levels in mixed effect meta-analysis models. First, we defined each treatment in each LTE according to common indices (scales or categories) of our identified EI practices and inputs (Table 1, Supplementary Table S1.2). Secondly, we estimated the mean yields and variances for the 'test crops' in each treatment in each LTE separately, using linear mixed models to account for the appropriate treatment and blocking structure. 'Test crops' were crops present in all treatments of an LTE: spring or winter wheat *(Triticum aestivum),* maize *(Zea mays),* oats *(Avena sativa),* barley *(Hordeum vulgare),* sugar beet *(Beta vulgaris)* or potatoes *(Solanum tuberosum).* We then calculated response ratios between the mean yields of each treatment within each LTE (henceforth termed 'yield ratios'). Finally, mixed effect meta-analysis models were applied to assess whether yield ratios responded consistently to particular EI practices or inputs across multiple cropping systems and locations, and whether the yield response to each EI practice or input was dependent on input levels and/or other parallel EI practices.

Separate meta-analysis models were applied for each of the three EI practices (CD, FC, and OM) and two inputs (NF and TI) to test the effect of changing one across different levels of the others. Unlike a standard meta-analysis approach that compares a 'response' treatment to a 'control' treatment, our meta-analysis models were constructed to compare multiple treatments, by specifying contrasts between various 'reference' and 'comparison' treatments in each LTE. Given our aim of exploring whether EI practices can increase yields for a given level of inputs or sustain yields while inputs are decreased, the contrast between a reference and comparison treatment always comprised either an increase or change in an EI practice or a reduction in an input (the nature or magnitude of which was described using moderator variables in the meta-analysis models). Each EI practice or input was the 'focal' variable in its own meta-analysis, and a 'context' variable in the meta-analysis for other EI practices and inputs (Table 1).

Using multiple models based on a common set of EI and input variables allowed us to robustly identify emergent overarching patterns in the yield response to different EI practices and inputs across the 30 LTEs. However, it should be noted that our yield ratio estimates for each specific combination of EI practices and inputs are representative of the subset of LTEs that tested those treatments, which determines the extent to which the findings are generalisable (not all treatments were replicated across a range of crop types, soil types, and climates). The confidence intervals in Figures 1-4 are important to indicate which treatment combinations are underpinned by more or less evidence: wide confidence intervals indicate estimates for treatments that were tested in fewer LTEs and/or where treatment effects were either inconsistent between replicates or years within each LTE, and/or inconsistent between LTEs (or all of the above).

Supporting information for the results presented in this paper is provided in the Supplementary Materials: Part 1 details each LTE and the treatments therein, Part 2 explains the use of Simpson's index as a metric for cropping system diversity, and Part 3 provides information on the meta-analysis models to support the interpretation of each result and the extent to which it is generalisable (including model selection metrics, significance tests for parameters, and tables and forest plots illustrating the contribution of each LTE to each model and treatment estimate).

Crop diversification (CD) and fertility crops (FC)

Both diversifying from a monoculture and adding fertility crops to an arable rotation usually increased test crop yields (Figure 1). However, NF interacted with legumes to moderate the effect of diversification. For CD, diversification with legumes resulted in a yield increase when NF was low (<100 kg N ha^{-1}) but not when NF was high (>100 kg ha^{-1}), while diversification with non-legumes resulted in a greater yield increase under high NF than low NF (Figure 1a). These results suggest different ecological functions are provided by different crop types: legumes contributed to test crop yields via biological nitrogen fixation when NF was low[33], whereas non-legumes likely contributed via regulation of weeds, pests and diseases (which becomes more important at high NF[34]).

FC also generally had the highest benefit when leguminous FC were added under low NF (Figure 1b). Under high NF, no FC crops significantly increased yields, and we observed a yield decrease when grain legumes were added to a ploughed arable rotation under high N, suggesting a possible antagonism between applied N and legumes in this context (although as only two European LTEs tested FC in this context, Supplementary Figure 3.2, the results may not be generalisable). There was also little benefit of adding annual FC to a rotation that already contained legumes and received NF, indicating that the additional biological nitrogen fixation function was redundant in this context. However, multi-annual FC, whether leguminous or grass leys, had benefits under low NF regardless of whether legumes were already present. This suggests leys provide additional functionality compared with annual FC, although leys still did not significantly increase yields under high NF.

Organic matter management (OM)

OM amendments were usually beneficial to long-term yields (Figure 2), although adding manure was associated with a larger yield increase than adding plant-based OM. It is possible this difference was due to greater quantities of manure compared with plant-based OM applied on average across our LTEs, or a higher nutrient content in the manure. Our assessment of the effects of different amendments was limited to the simple qualitative distinction of whether they were of plant or animal origin (Table 1), because the quantity, nutrient content, and type (e.g., plant species/fresh/composted) of OM varied too much between LTEs to explore more detailed effects in this study. We recommend further research using LTEs with more consistent OM treatments to compare different amendments more rigorously.

The yield benefit of OM amendments was greater under low NF and in systems without legumes (Figure 2) suggesting that nutrient input was an important contribution of OM to yields. In combination with our finding that diversifying with legumes is more beneficial under low NF (Figure 1), this suggests that N supply is an important aspect of the contribution of both legumes and OM to yields, but that multiple sources of N are not necessarily more effective than a single source. However, unlike legumes, adding OM under high NF does still have a small additional yield benefit, perhaps related to other nutrients such as phosphorus and potassium and their rate of release[35], or to increasing soil carbon and improving soil structure[36].

We did not observe a significant effect of retaining rather than removing crop residues on crop yields (Supplementary Table S3.3). This contradicts other research suggesting that residues can benefit yields through suppressing weeds, supporting beneficial biodiversity, improving water infiltration and conserving soil moisture[37]. Possibly residues have very site-specific effects, relating to residue type[38] and local pedo-climatic conditions[39,40], so our analysis could not identify a consistent dataset-wide effect (Supplementary Figure S3.3 shows that adding residues had small positive effects for come crop types in some LTEs and small negative effects in others). Surface residues under reduced tillage could also have different effects on soil properties and yields compared to ploughed-in residues[22], but we could not assess this interaction as only one LTE in our collection tested both residues and tillage together ('NTR' at SLU, Supplementary Tables S1.1, S1.2 and S3.8).

Reducing tillage and nitrogen fertiliser inputs

Of the two anthropogenic inputs investigated in this study, we found that reducing NF had strong negative effects on yield, while reducing TI had, at most, a slight negative effect. This suggests reducing TI may be an easy win to gain some environmental benefits (and potentially also climate resilience benefits[41]) while sustaining yields at or near current levels. Viewed from the opposite perspective, it also suggests that increasing TI does not substantially increase yields.

Our results on TI need cautious interpretation. Our null model, which tested only the effect of 'reducing tillage' without specifying which tillage practices were compared, indicated a mean yield ratio of 0.96 (a 4% decrease) on average across our dataset that was significantly different from zero ($Z = -2.097$, $P < 0.05$). The null model also suggested that no heterogeneity remained to be explained by the TI or context variables (QE $P>0.05$, Supplementary Table S3.1), although, when tillage type was included in the model, it did explain some heterogeneity (Supplementary Table S3.1, Supplementary Table S3.3). Taken together, these models indicate that the change in yield relating to TI is small compared to overall yield variability in the dataset, but there is some (inconclusive) evidence that different changes in TI result in different yield outcomes. For example, basins may have resulted in slightly higher yields than more intensive tillage, while shifting to no-till or zero-till may have slightly reduced yields on average (Figure 3), as has been observed in other studies[40,42].

In contrast to reducing TI, reducing NF had a strong but context-specific effect on yields (Supplementary Table S3.3). Our results show the standard asymptotic N response curve typically seen in cereal crops, but in reverse, because we tested the effect of incrementally reducing NF on yield ratios (Figure 4). This curve is modified by different context variables representing different EI practices. OM amendments and legumes both prevent the end of the curve where all N is removed from falling as low as it would in the absence of EI practices, showing that OM and legumes partly support yields when N fertiliser is low or absent. Manure had the strongest effect in this regard: if a system received manure applications, then most or all of the N fertiliser could be removed without seeing a yield reduction. In this study, reduced tillage may also have mitigated the effects of N removal (Figure 4), but too few LTEs tested different NF levels under reduced tillage to be certain, and other studies have suggested the opposite effect[42].

Overall, our results suggest an optimal level of NF that differs between contexts, but that is generally lower in the presence of EI practices. On average across all LTEs in our study, optimal NF was around 100 kg N ha^{-1}; Figure 4 demonstrates that reducing NF to this amount from higher NF rates did not reduce yields. Slightly more N could be removed without reducing yields if legumes were present, and more still if OM was present (especially manure), suggesting a lower optimal NF alongside these practices. Optimal N will also vary between different crops, climates, and soils (the 100 kg ha^{-1} figure given here is an average for our specific dataset and is not generalisable).

EI and inputs are substitutive or additive depending on function

A key finding of our study is that all EI practices assessed (CD, FC, and OM) increased long-term yields in most contexts, but the effects of EI practices and NF input on yields were partially substitutive: the benefits of EI practices were generally reduced at higher NF, and the requirement for NF was reduced when EI practices were employed. This indicates that N supply explains much of the contribution of the studied EI practices to crop yields. When crop demand for N is already met through fertiliser, only a relatively small additive benefit of EI practices was observed – for example, small yield increases from some forms of crop diversification (Figure 1) and OM amendments (Figure 2) when NF was high. These additive benefits likely indicate functions unique to different EI practices, such as ‘break crop’ functions of diversification[20], or nutrient cycling and soil structure improvements resulting from OM amendments[35,36]. These effects of different EI practices in different NF contexts are summarised in Figure 5. When NF is low (top panel), most EI practices increase yields whether they are applied separately or in combination, but especially if these EI practices have an N provisioning function (adding legumes or OM). In contrast, when NF is high (lower panel), then only EI practices that have functions distinct from N provisioning can increase yields.

In contrast to NF, tillage did not have a strong interaction with the EI practices, indicating that farmers may be able to make decisions about tillage and EI practices independently. We found the effect of reducing tillage to be small relative to the background variance in yield differences, but possibly slightly negative. This may not, however, be consistent among all forms of reduced tillage (Figure 3), and may also be influenced by environmental factors not assessed in this study – for instance, Sun et al[43] and Pittelkow et al[40,42] observed greater benefits in warmer, drier climates (suggesting the optimal TI for yield must balance a clean seedbed with soil water conservation). Furthermore, a small yield decrease may be acceptable in cases where reduced tillage offers non-yield benefits; either economic in terms of reduced fuel or labour costs, or environmental in terms of decreased soil erosion, increased water infiltration, or carbon sequestration[44].

Combining different EI practices together was more likely to result in positive effects than combining EI practices with anthropogenic inputs. The effect of diversification did not depend on whether OM was applied or not, indicating an additive benefit, while the effect of adding OM to diversified systems without legumes could be greater than the effect of adding OM to a monoculture, suggesting a possible synergistic effect. However, Figure 4 indicates that the absolute yield of systems containing combinations of only EI practices does tend to be lower than systems containing combinations of EI practices and moderate NF doses (compare yield ratios in the lower two rows where all NF is removed to where only some NF is removed). Thus, using EI in combination with some NF may best reduce the trade-off between input use and the land required to produce a given yield.

EI as a pathway to sustainable agriculture

In practical terms, a substitutive relationship between EI practices and N fertiliser means there is potential to: 1) use EI to increase yields when NF availability is low, 2) use EI to sustain a given yield while reducing NF levels, or 3) use EI to reduce the NF required to increase yields. However, combining high levels of NF with most EI practices does not increase yields. We also observed that antagonistic interactions between EI practices and high NF are possible; in particular, diversifying a highly fertilised system with legume crops may risk a yield decrease.

Widespread uptake of EI practices could therefore contribute to a more equitable global distribution of fertiliser. Currently, average NF rates in Africa are a small fraction of those in Europe, with smallholders in particular using much less than their fair share[45]. Both Foley et al[6] and Springmann et al[7] suggest that if fertiliser use is reduced where it is currently high, then fertiliser use could be increased where it is currently low, without exceeding planetary boundaries. EI practices could support this redistribution through sustaining yields while reducing fertiliser in current high-input high-yielding systems, and by enhancing yields in combination with moderate fertiliser inputs in currently low-yielding systems.

Future assessments of EI should include a wider analysis of farming systems and externalities. By focusing only on test crop yields, our study has not attempted to quantify implications for overall nutritional value or farm profitability. Currently it is difficult to use LTEs to assess whole-system performance, as too few LTEs rigorously measure yields of diverse crop types, nor do many collect measures of ecological function and socioeconomic outcomes. EI can have benefits beyond yield, by reducing the environmental and economic input costs to achieve a given yield[10,44]. Diversifying with legumes can increase profits and decrease pollution potential, by both increasing yields and reducing the fertiliser requirement of the whole rotation (assuming little or no fertiliser is applied to the legumes, and that fertilisation of subsequent crops is reduced), while also providing an additional potentially high value, protein-rich product[19]. Crop diversity can confer resilience to weather variability[41,46], increase biodiversity[47], and suppress weeds, pests and pathogens[20].

However, some practices that increase yields via ecological function (and that are thus considered EI in this study) may not necessarily avoid environmental impacts. For example, manures and composts reconnect resource flows between crops and livestock, but both can cause nutrient leaching and greenhouse gas emissions and so may not be objectively more environmentally friendly than NF. If manures and composts are available as waste products, however, their use as amendments at least recycles the nutrients therein and avoids further impacts from new synthetic fertiliser creation and use[21].

Socioeconomic factors can also limit the adoption of EI practices by farmers. These factors can include a lack of markets and infrastructure that can receive diverse products at viable prices[19,48], and limited access to necessary resources including land[49], seed, and OM sources[31]. Upscaling EI practices will thus require policymakers and society to create a more conducive socioeconomic context. Nonetheless, our results demonstrate that EI could play an important role in the development of future sustainable farming systems. Agricultural researchers could help to advance EI by further investigating which practices work best together in which contexts, to provide priorities for farmers and policymakers. We recommend that future LTEs place the development of a robustly ecologically functioning agroecosystem at the heart of their design and then explore what level of inputs are necessary to optimise the performance of these systems. Such LTEs would assist progress toward sustainable agriculture that remains within safe planetary boundaries whilst meeting human needs for food, fuel, and fibres.

Methods

LTEs included in this study contained at least one crop diversity (CD), fertility crop (FC), organic matter (OM), nitrogen fertiliser (NF) or tillage treatment (TI), and were located in either Europe or Africa. We defined an LTE as an 'experiment assessing the effect of treatments over decadal timescales' and thus all LTEs included were at least ten years old, with the exception of two nine-year-old LTEs in sub-Saharan Africa included in order to increase representation of smallholder farming systems. This minimum age ensured that the mean yield estimates for each treatment were unlikely to be driven by unusual weather in just one or two years. Suitable LTEs were identified and contacted via the GLTEN (www.glten.org) and authors' personal research networks. All LTEs that we could contact, that agreed to share their data, and that fit our criteria, were included in this study.

The LTEs were located in England, Kenya, Malawi, Mozambique, the Netherlands, Nigeria, Scotland, South Africa, Sweden, Zambia, and Zimbabwe (see Supplementary Figure S1.1). More details on each LTE, including the crop types, number of replicates and the number of years of data included can be found in Supplementary Materials Part 1.

Data analysis overview

We used a three-step analysis procedure to jointly interrogate the 30 LTEs in our dataset:

1. We first described each treatment in each LTE in terms of common EI and input indices that represented all treatments in all LTEs on comparable scales or in comparable categories;
2. We then estimated the yield mean and variance for each treatment in each LTE using a different linear mixed model for each LTE to account for the appropriate experimental design;
3. We explored how differences in mean crop yields between treatments related to differences in the common EI and input indices using mixed effect meta-analysis models.

The common EI and input indices (step 1) are described in Table 1, and the classification of each treatment in each LTE according to the common indices is detailed in Supplementary Table S1.2. The procedures for the individual mixed models (step 2) and the meta-analysis models (step 3) are described in the subsequent Methods sections, with supporting information for the meta-analyses provided in Supplementary Materials Part 3.

Together, these three steps comprised an efficient method to assess yield responses to comparable treatments across multiple cropping systems and locations. The meta-analysis approach allowed us to directly assess the size of the yield response ratio, and to identify the influence of moderating variables (different EI practice and input indices) on the size of the yield response ratio. Using mixed effects meta-analysis models helped to address limitations imposed by the number of LTEs that were available to include in our study: these models incorporate information about differences between treatments within LTEs, but also about differences between LTEs within shared treatments. Thus, when estimating treatment effects for rare treatment combinations (that may only occur in one or two LTEs), the meta-analysis uses information on the reliability of each LTE to inform the measures of certainty (confidence intervals and P-values) associated with each estimate. The models estimate treatment combinations with higher certainty if they are a) tested in LTEs that have limited within-LTE variation, b) tested in LTEs that have consistent effects with other LTEs included in the meta-analysis model, and/or c) tested in a greater number of LTEs.

Individual LTE models

To estimate yield means and variances for each treatment in each LTE, a separate linear mixed model was constructed for each LTE. Models were fitted in R version 4.0.2 using function *lmer* in package *lme4*[50]. All models followed the formula:

yield ~ treatment + (blocking structure) + (year)

where 'treatment' was a factor with each of the LTE's distinct treatments and/or treatment combinations as a different factor level. For example, if an LTE had three treatments consisting of a) a ploughed monoculture, b) a ploughed rotation and c) a no-till rotation, then the treatment factor for this model had three levels (a, b and c). Treatment was included as a fixed effect while the physical blocking structure and year were included as random terms.

Blocking structures were specified as appropriate for each LTE to account for the repeated crops grown in the same plot in multiple years (for example, sub-plot nested in main plot nested in block for a split-plot design). A random term for year was included as a factor to allow for variation between years and over time to be partitioned out, including if more recent years tended to have higher yields than past years, or vice versa. We did not account for additional temporal correlations between yields from the same plot in different years: in rotations of annual crops, the yield of a crop in one year is not strongly influenced by the yield of the same crop in previous years, with variation in weather likely to have a dominant impact[51].

Some initial models resulted in singular fits due to very low variance estimates for some random terms. Where this occurred the models were modified by including blocks as a fixed effect; this is often recommended for random terms with few levels (e.g. three blocks in an experiment) and does not change the model estimates and variances for each treatment. Average mean yields for each treatment across all blocks were estimated and their standard errors were calculated based on the pooled between-plot variability after allowing for any fixed block effects. If including blocks as a fixed effect did not suffice to avoid singularity, then we reduced the complexity of the random model by removing highly nested terms such as sub-plots nested within plots within blocks for which variances were estimated to be zero, or very close to zero[52].

The models included a weighting term to allow for the fact that the variance in the yields tended to increase as mean yield increased. Weights were obtained by running an unweighted model, obtaining the fitted values for each datapoint, and then including a weight of 1/(fitted value) in a second otherwise identical model. Plots of residuals were inspected to ensure this weighting was adequate to meet the assumption of homoscedasticity. Weights were not used for two models where the weights led to a non-convergence or singular result, and a plot of residuals indicated that weights were not needed to achieve homoscedasticity.

Where multiple test crops appeared in an LTE, a separate model was fitted for yields from each test crop. Data from all years after 1970 in which the test crop was grown were included. Years before 1970 were excluded to avoid introducing variability related to historical crop protection practices and crop cultivars (all LTEs had stopped using long straw cereal varieties by 1970). For LTEs that had more than one cropping season in a year (the four IITA LTEs in Kenya) then both seasons were included, and the random term for 'year' in the model was substituted by 'season.year' (each season in each year was treated as a separate event). Where crop failures occurred in some treatments but not others, these were included as zero yields for those treatments to capture the treatment-related variability. However, we made an exception for the complete crop failures that occurred across all treatments in all five South African LTEs in 2018, due to strong winds just prior to harvest dislodging the grain. For these LTEs in this year, yields were estimated based samples from small plots shortly before harvest. This provided information on treatment differences, whereas had we including the whole year as zero yields would have added noise to the mean yield estimates without providing information on treatment effects.

Data from plots in LTEs for treatments that were not relevant to this study were not included in these analyses, and the blocking structures suitably modified. Excluded treatments were those that received suboptimal levels of P and K in Rothamsted's Broadbalk, SRUC's Old Rotation, and the four SOM LTEs managed by IITA and ETH, and all 'historical manure' plots in Rothamsted's Woburn[53] (see Supplementary Table S1.1 for details of each LTE). These exclusions may have slightly inflated the variances associated with treatment estimates from these experiments due to not including all information about between-plot variability across the whole experiment, and this may have thus slightly increased Type II error rates (the probability of not detecting a true difference between treatments), but substantially streamlined the data collation process from these LTEs.

Where experiments underwent substantial changes that resulted in treatments being classified differently according to the common variables (e.g. transitions from long leys to short leys, or rotations with or without legumes), these were considered to be different treatments, and data from transition periods was excluded in order to only use data from established cropping systems. So for example, if an LTE began in 1980 but underwent substantial changes phased in between 1998 and 2002 (e.g. a change to the crop rotation or fertiliser treatments), an LTE could have treatments A, B and C from 1980 – 1998 and treatments D, E and F (on the same plots) after 2002, while the years 1998-2002 were discarded. These changes are detailed for each LTE in Supplementary Table S1.2.

Multi-LTE meta-analyses

The treatment means and their standard errors were extracted from the individual LTE models using function *emmeans* in package *emmeans*[54], then were summarised and collated into a single large dataset containing all pairwise contrasts of treatment combinations within each LTE (no pairs between LTEs). Each pair was labelled with the appropriate EI and input variables describing the 'reference' treatment and 'comparison' treatment, and with the LTE and crop type. Separate meta-analyses were conducted on this collated dataset to explore the effect of each EI and input focal variable in turn, accounting for the context variables in which the treatment contrasts occurred.

Each meta-analysis model was fitted to the subset of the data that included only the treatment contrasts relevant to the specific hypothesis for that focal variable (Table 1). Information on the treatment comparisons included from each LTE can be found in Supplementary Tables S3.4-10. It was not possible to fit a single meta-analysis model including all five EI and input forms as detailed variables, because there was insufficient replication of treatment combinations across the different LTEs.

All meta-analysis models were fitted in R (version 4.0.2) using package *metafor*[55]*.* Initially, the *escalc* function was used to calculate log response ratios and the associated variances for each treatment contrast in the combined dataset. In our study, the response is always crop yield and thus we term the response ratio the 'yield ratio' for clarity. These log yield ratios (weighted by their associated variances) formed the responses in a mixed effects meta-analysis model fitted using the function *rma.mv.* Both the focal and context variables were specified as moderators with fixed effects, while the LTE from which each treatment contrast originated was specified as a random effect to account for potential reduced independence among treatment contrasts from the same LTE. Where multiple test crops were present within an LTE, crop type was included as an additional random effect nested within LTE (not enough LTEs tested the same crops to include crop type as a fixed effect, or even a crossed random effect, which might have allowed more heterogeneity to be accounted for).

To identify the most appropriate meta-analysis model for each focal variable, several models of different complexities were fitted. This model selection process addressed the questions of 1) whether yield was affected by any directional change in the focal variable affected yield (null model, without any moderators), 2) whether the size of the specific change in the focal variable was important (base model, focal variable moderators only), 3) whether the effect of the focal variable depended on each of the context variables (intermediate model) and 4, whether the effects of different context variables on the effect of the focal variable on yield varied with the levels of the other context variables (full model) (see Supplementary Table S3.1). This approach did not consider all possible models but did allow assessment of the relative importance of different levels of complexity of the combinations of moderators. The best model was selected using the Akaike Information Criterion (AIC), and the QM and QE test statistics. The AIC describes 'goodness of fit' (heterogeneity explained by the model penalised by the complexity of the model), while the QM and QE test statistics assess the level of heterogeneity explained by the moderators included in the model, and the level of residual heterogeneity, respectively. The QM and QE test statistics are compared with the critical values of the appropriate chi-square distributions to calculate associated P-values.

A final model selection step was performed for the two continuous focal variables, crop diversity (Simpson's index) and nitrogen fertiliser reduction (reference nitrogen levels and proportion by which nitrogen differed between treatments). Initial models were fitted with second-order polynomials for each continuous variable, then equivalent models with only first-order polynomials. Models containing second-order polynomial terms were selected if these were significant as either main effects or interactions, and where removing the second-order polynomial terms increased the model AIC (Supplementary Table S3.2). The second-order polynomials were not intended to precisely describe the shape of the response curve to these focal variables, but simply to allow the model to identify if a curved relationship better described the data than a linear relationship.

When the model with the best fitting level of complexity had been identified, a QM test was conducted on each moderator main effect and interaction term included in the model to identify those which significantly influenced the mean yield ratio between treatments (Supplementary Table S3.3). The *anova.rma* function in *metafor* with the *btt* argument was used to specify each main effect or interaction separately. The importance of each term was assessed by comparing the QM test statistic with the critical values of the appropriate chi-square distribution. These tests assess the marginal contribution of each term to explain the heterogeneity in the response variable, while allowing for the effects of all other terms included in the model.

Meta-analyses that include multiple comparisons with a common control (or reference treatment, in the context of this study) can suffer bias due to a lack of independence between contrasts. However, this would have been at most a minor issue in this study, for the following reasons: 1) we extracted the contrasts directly from the full analysis of each LTE and have already accounted for any design-related non-independence[56]; 2) we included LTE as a random term in the meta-analyses models, accounting for that fact that yield ratios within the same LTE (and thus more likely to have a common reference treatment) are more related than yield ratios from different LTEs; and 3) all contrasts with common reference treatments in the intermediate and full models have different values of the moderator variables describing the comparison treatments, avoiding bias because these yield ratios are not pooled together[55]. A small amount of non-independence would not have been accounted for in the null models, and not always fully in the base models, due to the use of fewer moderators. This would not however have affected the model selection process (which is based on how well the moderators describe variation in the yield ratio), and so the only possible influence of non-independence would have occurred in our TI model. It may have slightly biased the overall estimate of the mean effect on yield of reducing tillage. However, as we emphasise in the main text, no strong conclusions should be drawn from this estimate anyway, given that the QE and QM values (Supplementary Table S3.3) indicate only a small effect of tillage relative to background variability in yields among the LTEs.

To plot the results for significant variables in each final meta-analysis model (Figures 1-4), the *predict.rma* function was used to calculate predictions and confidence intervals, interpolated within the range of each focal variable at each level of the context variables in the combined LTE dataset. Care was taken not to present predictions extrapolating beyond the range of the variables in the combined LTE dataset, given the use of polynomials for the CD and NF terms included in the models. Plots were constructed in package *ggplot2*[57].

Forest plots were also constructed to illustrate the contribution of each LTE to the meta-analysis estimates (Supplementary Figures S3.1-5). These show Best Linear Unbiased Predictions (BLUPS) that combine the fixed effect and random effect estimates for each LTE from the selected meta-analysis model for each EI theme (Supplementary Table S3.3). Essentially, these are the mean yield ratio estimates from each crop type in each LTE for a given treatment comparison. The plots also show standard errors for these mean yield ratio estimates, as estimates with a larger associated variance are given less weight in the meta-analysis model.

Classifying EI and input interactions

A key aim of this paper was to explore the interaction effects amongst the different EI practices and inputs to identify optimal combinations that maximised yield while minimising input use. We therefore classified interactions according to the following three established definitions[58] (in which A and B represent different EI practices or inputs): **antagonistic,** where the combined effect of A and B is less than the effect of either A or B alone; **additive,** where the combined effect of A and B is equal to the sum of the separate effects of A and B; and **synergistic,** where the combined effect of A and B is greater than the sum of the separate effects of A and B. We also included a fourth class of interaction that is less common in ecological and agricultural literature, **substitutive,** which we define as an interaction where the combined effect of A and B is the same as the maximum effect of either A or B alone, so that when A is reduced, the effect of B increases, and vice versa.

Acknowledgements

The authors would like to thank everyone who has been involved in designing, maintaining, funding, and collecting and managing data from all LTEs included in this study. We are grateful to the Global Long Term Experiment Network (GLTEN), https://glten.org/, funded by the Thirty Percy Foundation for providing meta-data on the LTEs. The Rothamsted Long-term Experiments National Capability (LTE-NC) is supported by the UK BBSRC (BBS/E/C/000J0300) and the Lawes Agricultural Trust. LTEs belonging to SRUC are supported through Scottish Government RESAS Strategic Research Programme. CM, JS, AMe and LC were supported by the 'GLTEN-Africa' project (BB/R020663/1) funded by the Global Challenge Research Fund programme of the Biotechnology and Biological Sciences Research Council (BBSRC), and JS and AMe also by the BBSRC Soils to Nutrition project (BBS/E/C/000I0320).

Author contributions

CM, JS, AMe and LC conceptualised this study. DB, AE, JH, WH, OJ, TK, JL, AMy, MN, GN, JSi, JSt, PS, CTh, CTo, FT, HV, RW, MW, and CW were involved in the management of and data collection from LTEs included. CM and AMe undertook the analysis with input from JS. CM wrote the manuscript with input from all authors.

Competing interests

The authors have no competing interests to declare.

Table 1: Detail on each EI practice and input investigated. The first column contains the hypotheses tested in each meta-analysis and the second column describes the treatment pairs used to calculate the yield ratios for each. All yield ratios compare a reference treatment, which is either a lower level of EI or a higher level of input, with a comparison treatment, which is a higher level of EI or a lower level of input. Bold text in the hypotheses in the first column describes the change between the reference and comparison treatments (tested in the null model), italicised text describes focal variables included as moderators (tested in the base model), while normal text describes the effects of the management context in which the EI practice or input reduction is implemented (tested in the intermediate and full models; Supplementary Table S3.1). Bold text in the second column indicates which characteristics of the reference and comparison treatments were included as focal moderator variables. The final column indicates the reduced number of levels used to describe the EI practice or input when it was a context variable in a meta-analysis with a different focal variable.

EI or input variable	Hypothesis	Focal variable(s) and treatment pairs	Levels when a context variable
CD: Crop diversification (EI practice)	**Shifting from a monoculture to a crop rotation or an intercrop will increase yields**, *and this increase will depend on how diverse the rotation or intercrop is, whether or not legumes are included*. It will also depend on the level of NF, OM, and TI in which this change in diversity is implemented.	Yield ratios were calculated between a monoculture* reference treatment with a Simpson's diversity index of 1, and a comparison treatment consisting of a rotation or an intercrop. The comparison treatment was characterised by whether it was a **rotation or an intercrop**, whether it **included legumes or not**, and by its **Simpson's index of diversity** (see Supplementary Materials Part 2). All yield ratios were calculated within levels of OM, TT and NF.	(1) Monoculture (2) Diverse with legumes (3) Diverse without legumes
FC: Fertility crops (EI practice)	**Adding a 'fertility crop' to an arable rotation will increase yields**. This could include adding a grain legume to an arable rotation without legumes, or adding a cover crop, forage crop or ley to an arable rotation with or without legumes. *The effect on yield of adding an FC crop will depend on whether the initial rotation contains legumes or not, the type of FC crop added, and whether or not the FC crop contains legumes*. It will also depend on the level of NF, OM, and TI in which this addition occurs.	Yield ratios were calculated between a **reference treatment** comprising an arable rotation either with or without grain legumes and a comparison treatment containing a **fertility crop**[†]: either an annual grain legume (to a rotation without legumes only), an annual service legume (cover crop, forage crop or hay crop), a multi-annual grass ley, or a multi-annual ley containing legumes (or a mix of legumes and others). We also considered whether the fertility crop was **grazed** by livestock physically present on the plots. All yield ratios were calculated within levels of OM, TT and NF.	(1) Diverse with legumes (2) Diverse without legumes
OM: Organic matter (EI practice)	**Adding organic matter, by either retaining crop residues or adding manure or plant materials (raw or composted), will increase yields**. *This increase will depend on the type of OM added*, and on the level of CD, NF, and TI in which that this addition occurs.	Yield ratios were calculated between treatment pairs that either described the addition of **organic matter amendments**, a change in the type of organic matter amendment, and/or a change in **crop residue management** from residue removal to residue retention. Organic matter amendments considered: none, plant-based, manure, or plant-based + manure. (No	(1) None (2) Plant-based OM added (3) Manure added

		plant-based amendments are living; all are cuttings, plant residues, compost, or biochar). All yield ratios were calculated within levels of CD, FC, TT, and NF.	
TI: Tillage intensity (input)	**Reducing tillage will increase yields***,* *and this increase will depend on the initial type of tillage and the type of tillage to which it is reduced*. It will also depend on the level of CD, NF, and OM.	Yield ratios were calculated for treatment pairs describing a reduction in tillage intensity, so the **reference treatment tillage type** always consisted of a more intensive practice than the **comparison treatment tillage type.** Tillage types were considered to rank in intensity in the following order: - Deep (15-25 cm) inversion tillage (e.g. mouldboard plough) - Ridge-furrow planting (soil dug over and shaped into ridges and furrows) - Deep (15-25cm) non-inversion tillage (e.g. subsoiling) - Shallow (5-10cm) non-inversion tillage (e.g. tine harrow) - Infrequent tillage (tillage less than once per year) - Basins (soil dug within confined areas to create planting basins; also known as zai) - No-till (no tillage but some soil disturbance caused by planting implement, e.g. tine openers, rip-line seeding) - Zero-till (no tillage and no soil disturbance caused by planting implements, e.g. disc openers, dibble sticks or jab planters). All yield ratios were calculated within levels of CD, FC, OM, and NF.	(1) Deep inversion tillage (2) Reduced tillage (includes ridge-furrows, non-inversion tillage, shallow tillage, basin, and infrequent tillage[‡]) (3) No till (includes no till and zero till)
NF: Nitrogen fertilisation (input)	**Reducing nitrogen fertilisation will affect yields**, *and this effect will depend on the initial amount of nitrogen applied and by how much it is reduced*. It will also depend on the level of CD, OM, and TI in which this reduction is implemented.	Yield ratios were calculated from treatment pairs describing a reduction in N fertilisation in the reference treatment and the comparison treatment, measured as the amount of **N fertiliser applied to the reference treatment** in kg N ha, and **proportion by which N was reduced** in the comparison treatment. All yield ratios were calculated within levels of CD, FC, OM, and TT.	(1) Zero N (2) Between 1 and 100 kg N ha^{-1} (3) More than 100 kg N ha^{-1}

Notes: *We define a monoculture as the same crop in a plot in every year (i.e. neither an intercrop nor a rotation). [†]A 'fertility crop' is defined as a crop with distinct properties from the test crops that can be added to a rotation with the aim of increasing yield through increasing the functional diversity of a system. We use this variable to explore the effects of adding nitrogen-fixing legumes, crops with different harvest procedures (cutting or grazing), and multi-annual leys, each of which could be expected to influence test crop yields via different mechanisms. [‡]That these practices are reduced relative to deep inversion tillage, not necessarily reduced in comparison to local conventional practices.

Figure Captions

Figure 1: Estimated mean yield ratios for **(a)** CD; diversifying from a monoculture to cropping systems with higher Simpson's indices, using diversity that includes or excludes grain legumes (panels), and **(b)** FC; adding fertility crops to a diverse arable system that either lacks legumes (left panel) or already contains grain legumes (right panel). In **(a)** the x axis does not extend below 1.5, as monocultures have a Simpson's index of 1 (and all treatment pairs in the model compared a reference monoculture to a comparison treatment with higher diversity). In both figures, the nitrogen context in which the diversification occurs is indicated by colours, while in **(b)** the tillage context is indicated by point shapes and grazing by point shading (these variables were not significant in the CD meta-analysis and so are not shown in **(a)**). The horizontal dashed line marks a yield ratio of 1, or no change. Error bars and ribbons indicate 95% confidence intervals for the mean yield ratio. The model results and forest plots of treatment contrasts underlying these predictions are shown in detail in Supplementary Materials Part 3.

Figure 2: Estimated mean yield ratios for different OM amendments (x axis) in different diversity contexts (columns), tillage contexts (point shapes) and nitrogen fertiliser contexts (colours). The labels on the x axis indicate the OM addition; 'add plant material' and 'add manure' are additions to systems currently not receiving any OM, while 'change plant material to manure' is the yield ratio between a system receiving plant-based OM and a system receiving manure, and 'add plant material to manure' is the yield ratio between a system receiving manure and a system receiving both types of OM addition. The horizontal dashed line marks a yield ratio of 1, or no change. Error bars indicate 95% confidence intervals for the mean yield ratio. The model results and forest plots of treatment contrasts underlying these predictions are shown in detail in Supplementary Materials Part 3.

Figure 3: Estimated mean yield ratios for shifting from a reference tillage treatment (x axis, and point colours for emphasis) to a reduced tillage comparison treatment (panels). For example, the yield ratio comparing deep inversion tillage with deep non-inversion tillage is just over 1 (leftmost panel), while most yield ratios between more intensive tillage systems and zero-till systems are negative, with the exception of shifting from a ridge-furrow system to zero-till (rightmost panel). Yield ratios were always calculated for *reducing* tillage, i.e. comparing the effect of shifting to a less intensive tillage treatment from a more intensive tillage treatment. The horizontal dashed line marks a yield ratio of 1, or no change. Error bars indicate 95% confidence intervals for the mean yield ratio. The model results and forest plots of treatment contrasts underlying these predictions are shown in detail in Supplementary Materials Part 3.

Figure 4: Estimated mean yield ratios for systems receiving a certain amount of NF (columns) compared to systems receiving a reduced quantity of fertiliser N (x axis; expressed as a proportion of the initial N kg ha^{-1}), in different diversity contexts (rows), tillage contexts (point or line shading) and organic matter contexts (colours). Points are shown instead of lines where only one or two N levels were present within a certain context, as a curved line cannot be reasonably estimated from only two points. The horizontal dashed line marks a yield ratio of 1, or no change. Error bars and ribbons indicate 95% confidence intervals for the mean yield ratio. The x axis does not extend to zero as all yield ratios in the model compared a higher reference amount of NF to a lower test amount of NF.

The model results and forest plots of treatment contrasts underlying these predictions are shown in detail in Supplementary Materials Part 3.

Figure 5: A summary of the EI practices and combinations thereof that increase yields (green arrows), have no effect on yields (yellow arrows), or may risk a yield decrease (orange arrow), when implemented in either a low NF context (top panel) or high NF context (lower panel). White boxes represent farming systems with specific EI practices, and moving from one white box to another along the direction of an arrow symbolises the addition of an EI practice to that system. Where arrows are not shown (e.g. adding a ley to a diverse system with OM), we did not have sufficient data to test this contrast in our study. CD or FC practices that include legumes typically resulted in yield increases under low NF, while in contrast, CD or FC practices that did not include legumes resulted in yield increases under high NF. Adding OM increased yields unless the system already contained legumes and received high NF. Tillage is not shown because we did not identify any clear and consistent interactions between tillage and different EI practices.

FIGURE 1

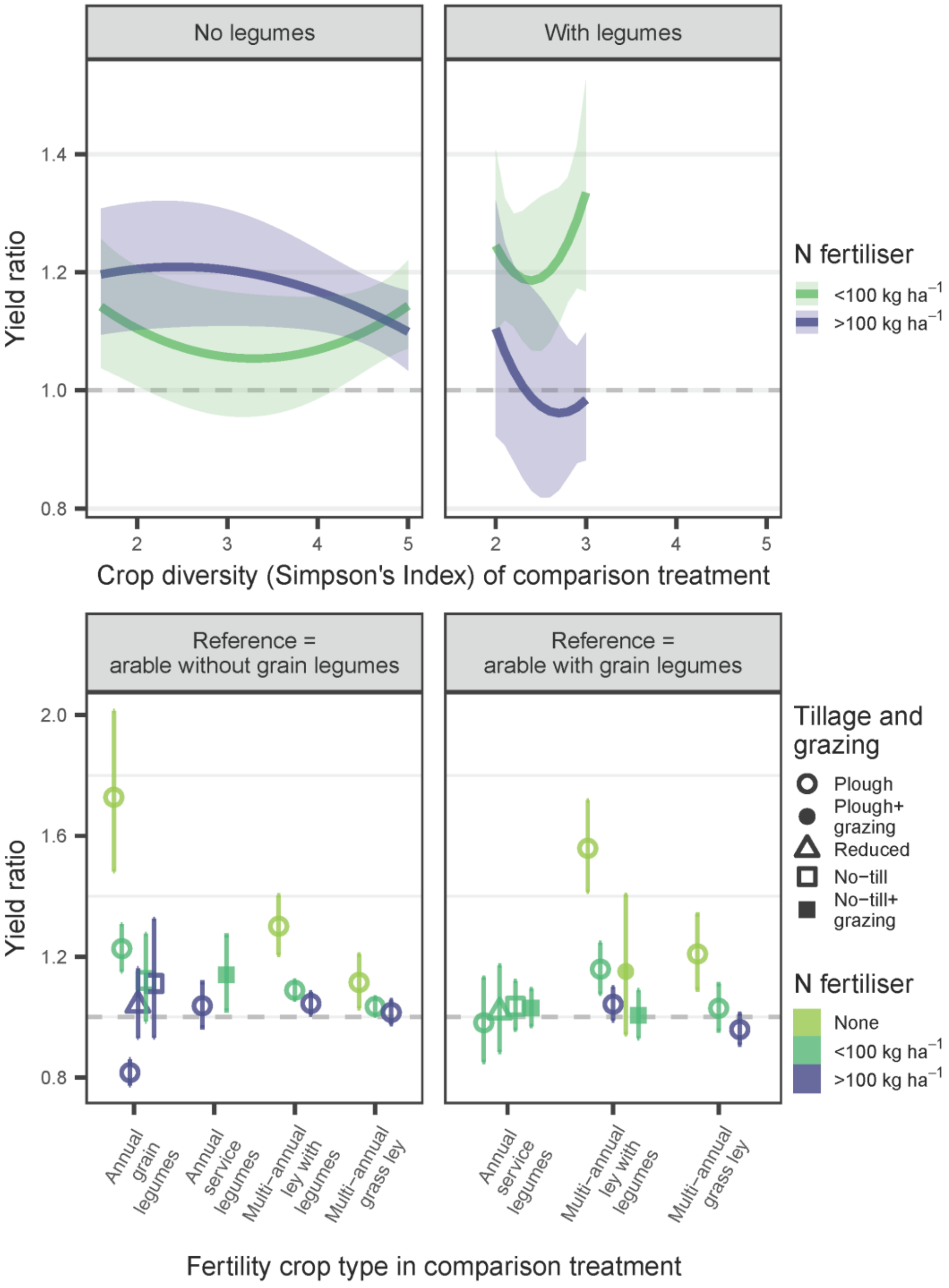

No legumes
With legumes
Yield ratio
1.4
1.2
1.0
0.8
2
3
4
5
Crop diversity (Simpson's Index) of comparison treatment
N fertiliser
<100 kg ha−1
>100 kg ha−1
Reference = arable without grain legumes
Reference = arable with grain legumes
Yield ratio
2.0
1.6
1.2
0.8
Tillage and grazing
Plough
Plough+ grazing
Reduced
No−till
No−till+ grazing
N fertiliser
None
<100 kg ha−1
>100 kg ha−1
Annual grain legumes
Annual service legumes
Multi−annual ley with legumes
Multi−annual grass ley
Annual service legumes
Multi−annual ley with legumes
Multi−annual grass ley
Fertility crop type in comparison treatment

FIGURE 2

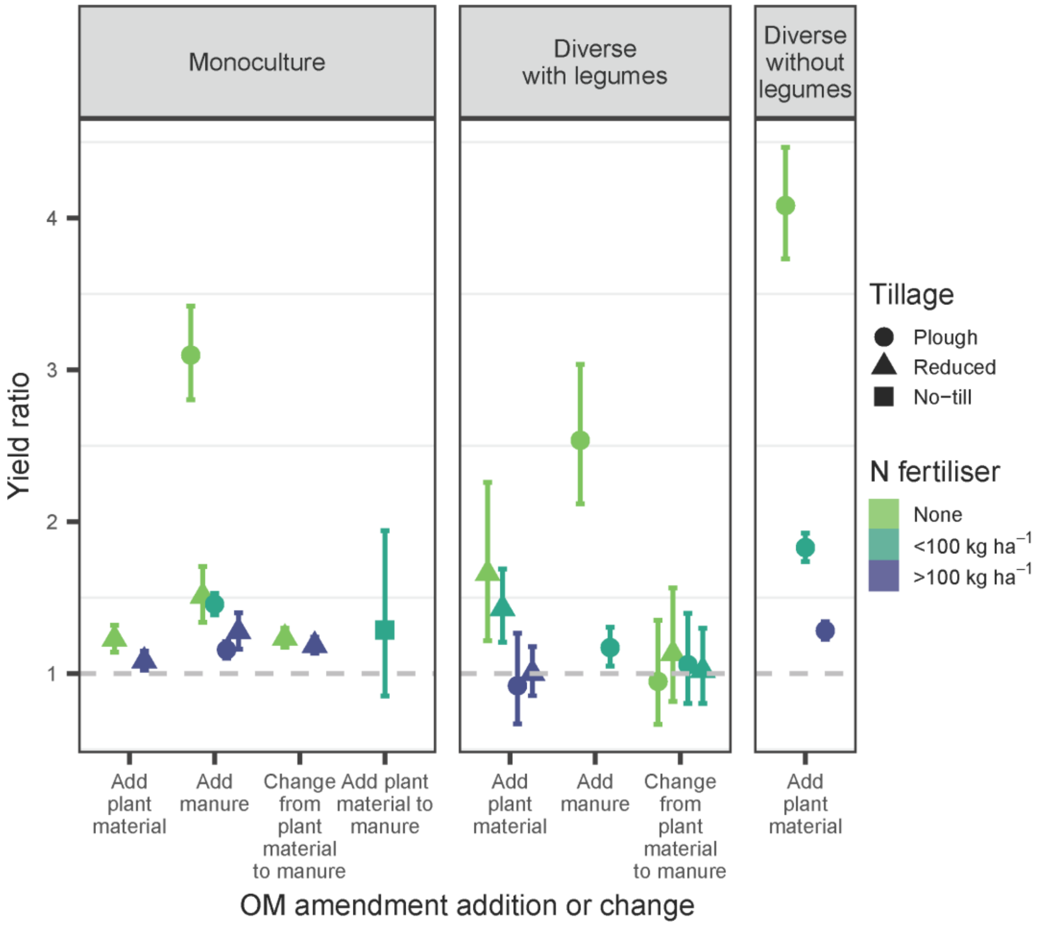

Monoculture
Diverse with legumes
Diverse without legumes
Yield ratio
4
3
2
1
Add plant material
Add manure
Change from plant material to manure
Add plant material to manure
Add plant material
Add manure
Change from plant material to manure
Add plant material
OM amendment addition or change
Tillage
Plough
Reduced
No−till
N fertiliser
None
<100 kg ha^{-1}
>100 kg ha^{-1}

FIGURE 3

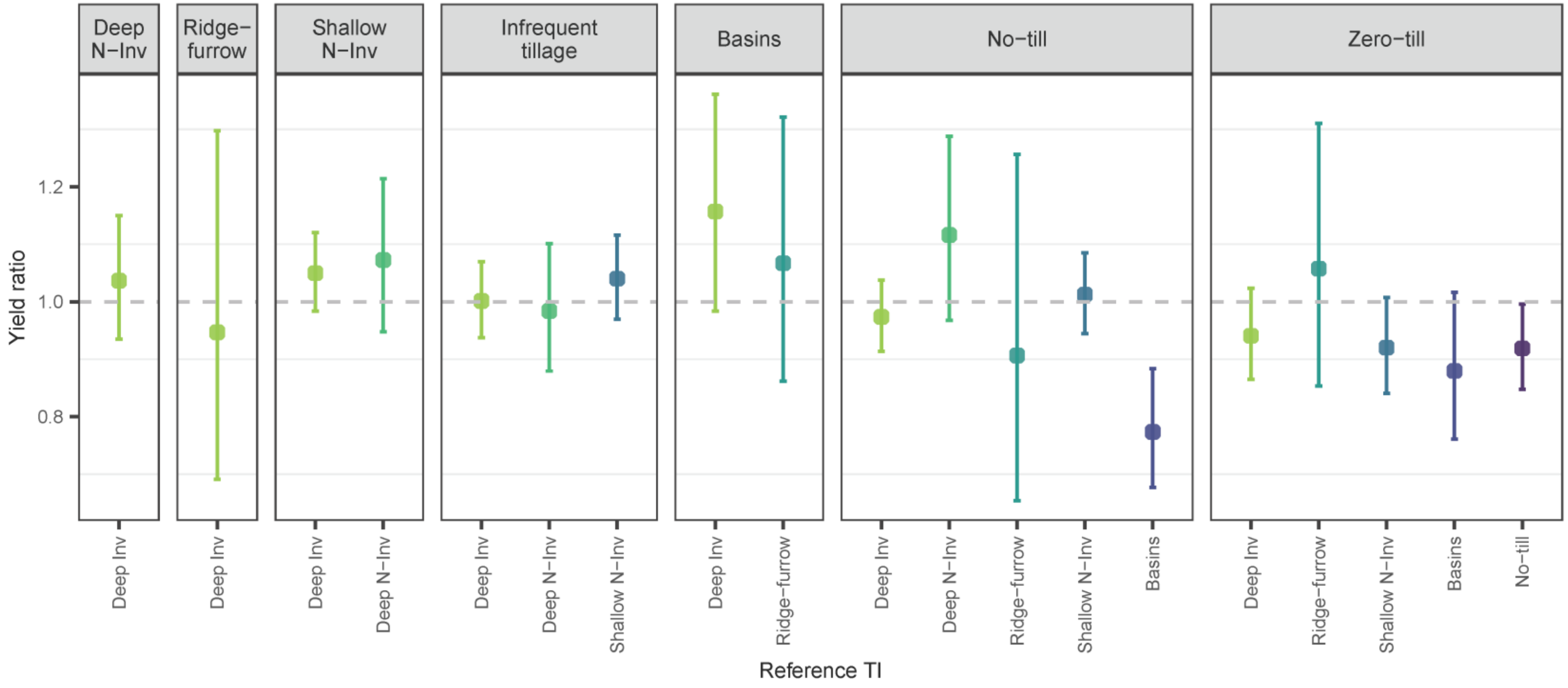

Deep N−Inv
Ridge−furrow
Shallow N−Inv
Infrequent tillage
Basins
No−till
Zero−till
Yield ratio
1.2
1.0
0.8
Deep Inv
Deep Inv
Deep Inv
Deep N−Inv
Deep Inv
Deep N−Inv
Shallow N−Inv
Deep Inv
Ridge−furrow
Deep Inv
Deep N−Inv
Ridge−furrow
Shallow N−Inv
Basins
Deep Inv
Ridge−furrow
Shallow N−Inv
Basins
No−till
Reference TI

FIGURE 4

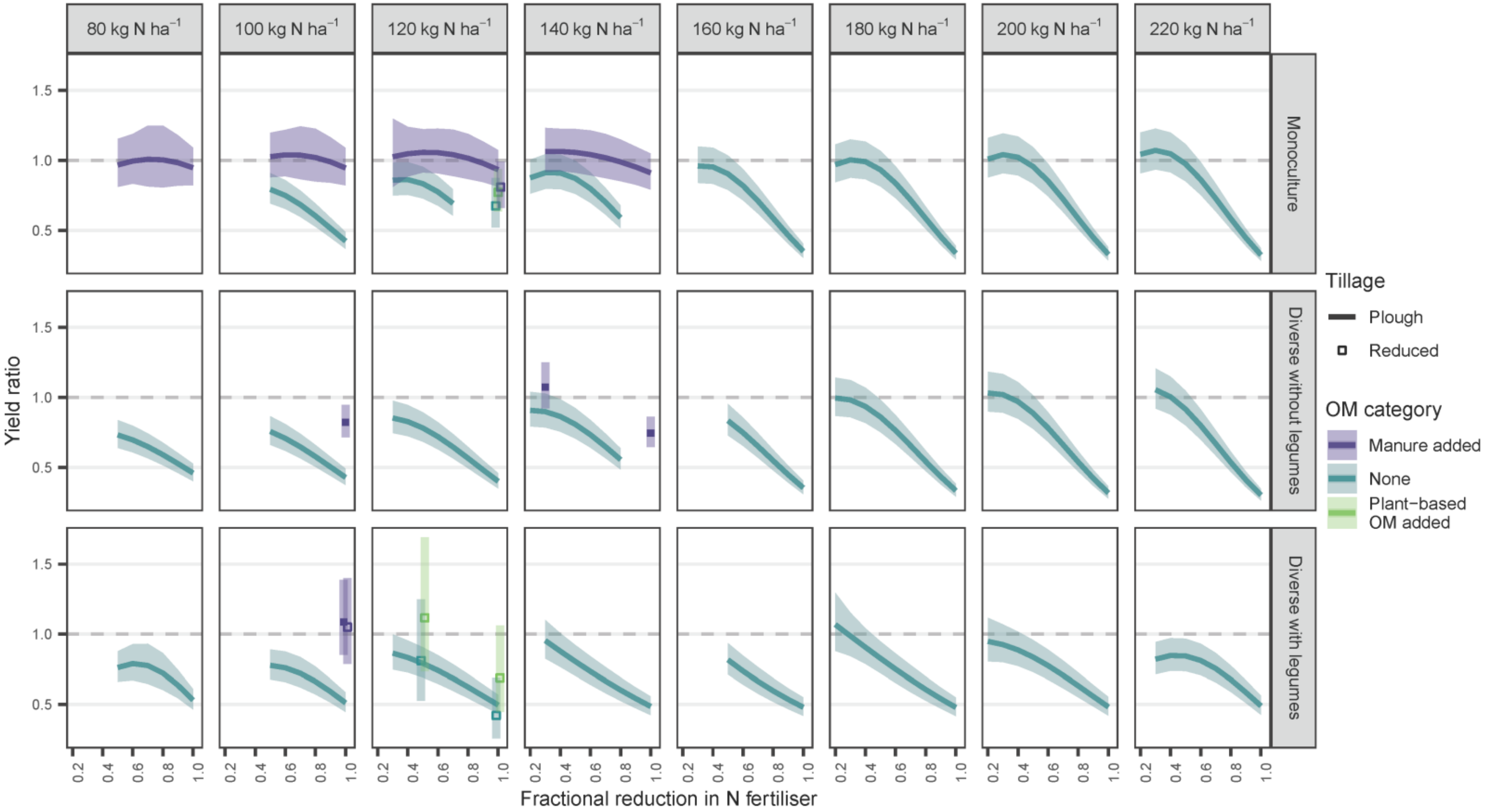
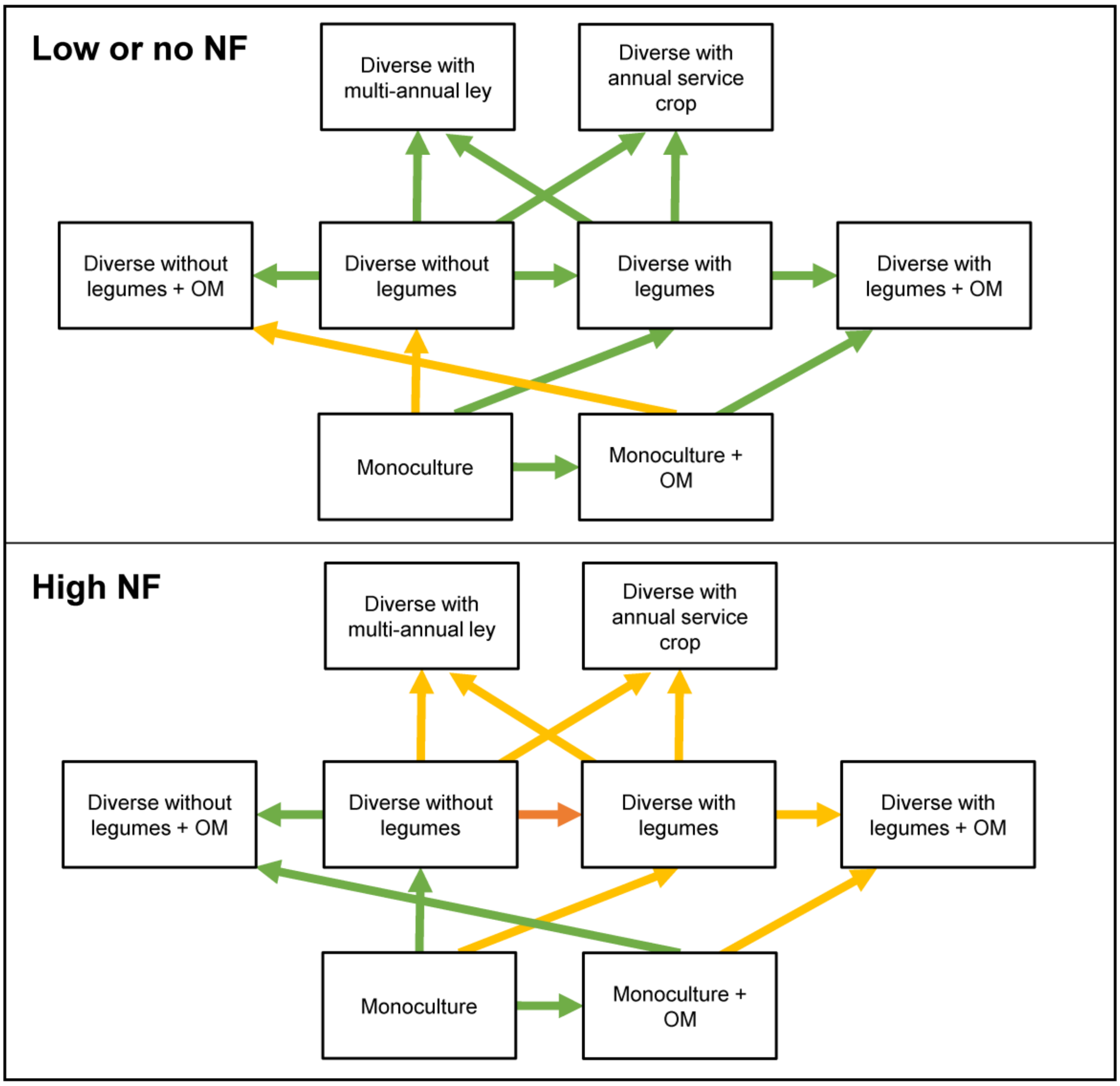

80 kg N ha^{-1}
100 kg N ha^{-1}
120 kg N ha^{-1}
140 kg N ha^{-1}
160 kg N ha^{-1}
180 kg N ha^{-1}
200 kg N ha^{-1}
220 kg N ha^{-1}
Monoculture
Diverse without legumes
Diverse with legumes
Yield ratio
0.5
1.0
1.5
0.2
0.4
0.6
0.8
1.0
Fractional reduction in N fertiliser
Tillage
Plough
Reduced
OM category
Manure added
None
Plant−based OM added

FIGURE 5


Low or no NF
Diverse with multi-annual ley
Diverse with annual service crop
Diverse without legumes + OM
Diverse without legumes
Diverse with legumes
Diverse with legumes + OM
Monoculture
Monoculture + OM
High NF
Diverse with multi-annual ley
Diverse with annual service crop
Diverse without legumes + OM
Diverse without legumes
Diverse with legumes
Diverse with legumes + OM
Monoculture
Monoculture + OM

Data availability

The datasets analysed during the current study are available from the authors on reasonable request. Please contact the corresponding author for assistance. Data from LTEs belonging to Rothamsted Research are available on reasonable request via the e-RA platform (www.era.rothamsted.ac.uk). We have refrained from depositing data into a public repository due to the need for guidance to correctly interpret LTE designs and datasets, and the need to ensure that the substantial investments by each institute in maintaining LTEs do not go unacknowledged when data is used.

Code availability

R scripts used in the analyses are also available from the corresponding author on reasonable request.

References

1. Rockström, J. *et al.* Sustainable intensification of agriculture for human prosperity and global sustainability. *Ambio* **46**, 4–17 (2017).
2. Steffen, W. *et al.* Planetary boundaries: Guiding human development on a changing planet. *Science (80-.).* **347**, 1259855 (2015).
3. Campbell, B. M. *et al.* Agriculture production as a major driver of the Earth system exceeding planetary boundaries. *Ecol. Soc.* **22**, 8 (2017).
4. Hazell, P. & Wood, S. Drivers of change in global agriculture. *Philos. Trans. R. Soc. B Biol. Sci.* **363**, 495–515 (2008).
5. Lehmann, P. *et al.* Complex responses of global insect pests to climate warming. *Front. Ecol. Environ.* **18**, 141–150 (2020).
6. Foley, J. A. *et al.* Solutions for a cultivated planet. *Nature* **478**, 337–342 (2011).
7. Springmann, M. *et al.* Options for keeping the food system within environmental limits. *Nature* **562**, 519–525 (2018).
8. Hunter, M. C., Smith, R. G., Schipanski, M. E., Atwood, L. W. & Mortensen, D. A. Agriculture in 2050: recalibrating targets for sustainable intensification. *Bioscience* **67**, 386–391 (2017).
9. Millenium Ecosystem Assessment. *Ecosystems and Human Well-being: Synthesis.* (2005) doi:10.5822/978-1-61091-484-0_1.
10. Bommarco, R., Kleijn, D. & Potts, S. G. Ecological intensification: harnessing ecosystem services for food security. *Trends Ecol. Evol.* **28**, 230–238 (2013).
11. Kleijn, D. *et al.* Ecological Intensification: Bridging the Gap between Science and Practice. *Trends Ecol. Evol.* **xx**, 1–13 (2018).
12. Pingali, P. L. Green revolution: Impacts, limits, and the path ahead. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 12302–12308 (2012).
13. Wezel, A. *et al.* Agroecology as a science, a movement and a practice. *Sustain. Agric.* **2**, 27–43 (2009).
14. Garnett, T. *et al.* Sustainable intensification in agriculture: Premises and policies. *Science (80-.*

). **341**, 33–34 (2013).

15. Lipper, L. *et al.* Climate-smart agriculture for food security. *Nat. Clim. Chang.* **4**, 1068–1072 (2014).

16. Tittonell, P. Ecological intensification of agriculture — sustainable by nature. *Curr. Opin. Environ. Sustain.* **8**, 53–61 (2014).

17. Jenkinson, D. S. The impact of humans on the nitrogen cycle, with focus on temperate arable agriculture. *Plant Soil* **228**, 3–15 (2001).

18. Verheijen, F. G. A., Jones, R. J. A., Rickson, R. J. & Smith, C. J. Tolerable versus actual soil erosion rates in Europe. *Earth-Science Rev.* **94**, 23–38 (2009).

19. Peoples, M. B. *et al.* The Contributions of Legumes to Reducing the Environmental Risk of Agricultural Production. in *Agroecosystem Diversity: Reconciling Contemporary Agriculture and Environmental Quality* 123–142 (Academic Press, 2019). doi:10.1016/B978-0-12-811050-8.00008-X.

20. Storkey, J., Bruce, T., McMillan, V. & Neve, P. The future of sustainable crop protection relies on increased diversity of cropping systems and landscapes. in *Agroecosystem Diversity* (eds. Lemaire, G., Carvalho, P. C. de F., Kronberg, S. & Recous, S.) 199–209 (Academic Press, 2019). doi:https://doi.org/10.1016/B978-0-12-811050-8.00012-1.

21. Schröder, J. Revisiting the agronomic benefits of manure: A correct assessment and exploitation of its fertilizer value spares the environment. *Bioresour. Technol.* **96**, 253–261 (2005).

22. Mhlanga, B., Ercoli, L., Pellegrino, E., Onofri, A. & Thierfelder, C. The crucial role of mulch to enhance the stability and resilience of cropping systems in southern Africa. *Agron. Sustain. Dev.* **41**, 29–43 (2021).

23. Sandén, T. *et al.* European long-term field experiments: knowledge gained about alternative management practices. *Soil Use Manag.* **34**, 167–176 (2018).

24. Storkey, J. *et al.* The unique contribution of Rothamsted to ecological research at large temporal scales. *Adv. Ecol. Res.* **55**, 3–42 (2016).

25. Johnston, A. E. & Poulton, P. R. The importance of long-term experiments in agriculture: their management to ensure continued crop production and soil fertility; the Rothamsted experience. *Eur. J. Soil Sci.* **69**, 113–125 (2018).

26. Bowles, T. M. *et al.* Long-term evidence shows that crop-rotation diversification increases agricultural resilience to adverse growing conditions in North America. *One Earth* **2**, 1–10 (2020).

27. Marini, L. *et al.* Crop rotations sustain cereal yields under a changing climate. *Environ. Res. Lett.* **15**, 124011 (2020).

28. Lal, R. Carbon emission from farm operations. *Environ. Int.* **30**, 981–990 (2004).

29. Cordell, D., Drangert, J. O. & White, S. The story of phosphorus: Global food security and food for thought. *Glob. Environ. Chang.* **19**, 292–305 (2009).

30. Lechenet, M., Dessaint, F., Py, G., Makowski, D. & Munier-Jolain, N. Reducing pesticide use while preserving crop productivity and profitability on arable farms. *Nat. Plants* **3**, 17008 (2017).

31. Barrett, C. B. & Bevis, L. E. M. The self-reinforcing feedback between low soil fertility and chronic poverty. *Nat. Geosci.* **8**, 907–912 (2015).

32. Tittonell, P. & Giller, K. E. When yield gaps are poverty traps: the paradigm of ecological intensification in African smallholder agriculture. *F. Crop. Res.* **143**, 76–90 (2013).

33. Bedoussac, L. *et al.* Ecological principles underlying the increase of productivity achieved by cereal-grain legume intercrops in organic farming. A review. *Agron. Sustain. Dev.* **35**, 911–935 (2015).

34. Storkey, J., Mead, A., Addy, J. & MacDonald, A. J. Agricultural intensification and climate change have increased the threat from weeds. *Glob. Chang. Biol.* **27**, 2416–2425 (2021).

35. Vanlauwe, B. *et al.* Direct interactions between N fertilizer and organic matter: evidence from trials with 15 N labelled fertilizer. in *Integrated Plant Nutrient Management in sub-Saharan Africa: From Concept to Practice* (eds. Vanlauwe, B., Diels, J., Sanginga, N. & Merckx, R.) 173–184 (CABI, 2002).

36. Hijbeek, R. *et al.* Do organic inputs matter – a meta-analysis of additional yield effects for arable crops in Europe. *Plant Soil* **411**, 293–303 (2017).

37. Thierfelder, C. & Wall, P. C. Effects of conservation agriculture techniques on infiltration and soil water content in Zambia and Zimbabwe. *Soil Tillage Res.* **105**, 217–227 (2009).

38. Gentile, R., Vanlauwe, B., Chivenge, P. & Six, J. Interactive effects from combining fertilizer and organic residue inputs on nitrogen transformations. *Soil Biol. Biochem.* **40**, 2375–2384 (2008).

39. Mupangwa, W. *et al.* Maize yields from rotation and intercropping systems with different legumes under conservation agriculture in contrasting agro-ecologies. *Agric. Ecosyst. Environ.* **306**, 107170 (2021).

40. Pittelkow, C. M. *et al.* Productivity limits and potentials of the principles of conservation agriculture. *Nature* **517**, 365–368 (2015).

41. Steward, P. R. *et al.* The adaptive capacity of maize-based conservation agriculture systems to climate stress in tropical and subtropical environments: A meta-regression of yields. *Agric. Ecosyst. Environ.* **251**, 194–202 (2018).

42. Pittelkow, C. M. *et al.* When does no-till yield more? A global meta-analysis. *F. Crop. Res.* **183**, 156–168 (2015).

43. Sun, W. *et al.* Climate drives global soil carbon sequestration and crop yield changes under conservation agriculture. *Glob. Chang. Biol.* **In press**, (2020).

44. Kirkegaard, J. A. *et al.* Sense and nonsense in conservation agriculture: principles, pragmatism and productivity in Australian mixed farming systems. *Agric. Ecosyst. Environ.* **187**, 133–145 (2014).

45. Thierfelder, C. *et al.* Complementary practices supporting conservation agriculture in southern Africa. A review. *Agron. Sustain. Dev.* **38**, 16–37 (2018).

46. Bowles, T. M. *et al.* Long-term evidence shows that crop-rotation diversification increases agricultural resilience to adverse growing conditions in North America. *One Earth* **2**, 1–10 (2020).

47. Alignier, A. *et al.* Configurational crop heterogeneity increases within-field plant diversity. *J. Appl. Ecol.* **57**, 654–663 (2020).

48. Liebman, M. *et al.* Ecologically sustainable weed management: How do we get from proof-of-concept to adoption? *Ecol. Appl.* **26**, 1352–1369 (2016).

49. Giller, K. E. The Food Security Conundrum of sub-Saharan Africa. *Glob. Food Sec.* **26**, 100431 (2020).

50. Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

51. Addy, J. W. G., Ellis, R. H., Macdonald, A. J., Semenov, M. A. & Mead, A. Changes in agricultural climate in South-Eastern England from 1892 to 2016 and differences in cereal and permanent grassland yield. *Agric. For. Meteorol.* **308**–**309**, 108560 (2021).

52. Bates, D., Kliegl, R., Vasishth, S. & Baayen, H. Parsimonious Mixed Models [preprint]. *arXiv:1506.04967v2* (2018).

53. MacLaren, C., Glendining, M., Poulton, P., Macdonald, A., Clark, S. Woburn Ley-arable experiment: yields of wheat as first test crop, 1976-2018. *e-RA*. http://dx.doi.org/10.23637/wrn3-wheat7618-01 (2022).

54. Lenth, R. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.7.2. *https://CRAN.R-project.org/package=emmeans* (2020).

55. Viechtbauer, W. Conducting meta-analyses in R with the metafor. *J. Stat. Softw.* **36**, 1–48 (2010).

56. Lajeunesse, M. J. On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology* **92**, 2049–2055 (2011).

57. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

58. Crain, C. M., Kroeker, K. & Halpern, B. S. Interactive and cumulative effects of multiple human stressors in marine systems. *Ecol. Lett.* **11**, 1304–1315 (2008).