

Global crop production increase by soil organic carbon

Received: 17 April 2023

Accepted: 21 September 2023

Published online: 30 October 2023

 Check for updates

Yuqing Ma  ^{1,2,7}, Dominic Woolf  ^{2,3,4,7}, Mingsheng Fan  ¹✉, Lei Qiao  ⁵, Rong Li ⁶ & Johannes Lehmann  ^{2,3,4}✉

Soil organic carbon sequestration has been promoted to combat climate change while improving soil fertility. However, its quantitative contribution to crop productivity has proven elusive. Using data from 13,662 controlled field trials with 66,593 treatments across a broad range of soils, climates and management practices, we here show that yields increase with increased soil organic carbon, until no further increase ($p < 0.05$) occurs above mean optimum soil organic carbon of 43.2–43.9 g kg⁻¹ for maize, 12.7–13.4 g kg⁻¹ for wheat and 31.2–32.4 g kg⁻¹ for rice. Sequestering soil organic carbon is one-fifth as effective (that is, 80% less) as nitrogen fertilization for improving crop yield where soil management is optimized. By increasing soil organic carbon beyond current technology to optimum levels, global production of the three most important staple crops increases by 4.3% (sufficient to provide calories for 640 million people). However, currently available management practices would increase crop production by only 0.7% once other production constraints have already been addressed. Therefore, yield improvements under currently available technologies are unlikely to drive adoption of soil organic carbon sequestration globally, except in hot-spot regions where crop production benefits most, or unless novel practices that allow greater soil organic carbon sequestration beyond current limitations can further increase yields cost-effectively.

With global greenhouse gas (GHG) emissions still rising¹, the window of opportunity to avoid dangerous climate change through emissions reduction alone has passed². Carbon dioxide removal from the atmosphere is therefore also expected to be required², within a broader portfolio of mitigation measures that must simultaneously reduce emissions from energy and industry³. Sequestering photosynthetically fixed organic carbon in soil organic carbon (SOC) has been suggested as a potentially low-cost and highly scalable approach to carbon dioxide removal⁴. At a global scale, SOC contains about three times more carbon than atmospheric carbon dioxide (CO₂)⁵. Therefore, a small fractional increase in SOC

stocks could have a substantial impact on CO₂ concentrations⁶. With most of the potential to increase SOC located on agricultural soils, long-term adoption of SOC sequestration practices depends on buy-in from farmers^{6,7}. The promise of improved soil health leading to higher crop yields has been suggested as the most important driver of adoption because this would support both farm productivity and food security objectives⁷.

However, although it has long been hypothesized that higher SOC can increase crop yields^{8,9}, the causal impact of SOC on yield at a regional or global scale has still not been unequivocally demonstrated or quantified¹⁰. The causal impact of SOC on crop yield at large scales

¹State Key Laboratory of Nutrient Use and Management, College of Resources and Environmental Sciences, China Agricultural University, Beijing, China.

²Soil and Crop Sciences, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA. ³Cornell Institute for Digital Agriculture (CIDA), Cornell University, Ithaca, NY, USA. ⁴Cornell Atkinson Center for Sustainability, Cornell University, Ithaca, NY, USA. ⁵Key Laboratory of Arable Land Quality Monitoring and Evaluation, Ministry of Agriculture and Rural Affairs, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China. ⁶Cultivated Land Quality Monitoring and Protection Center, Ministry of Agriculture and Rural Affairs of the People's Republic of China, Beijing, China. ⁷These authors contributed equally: Yuqing Ma, Dominic Woolf. ✉e-mail: fanms@cau.edu.cn; CL273@cornell.edu

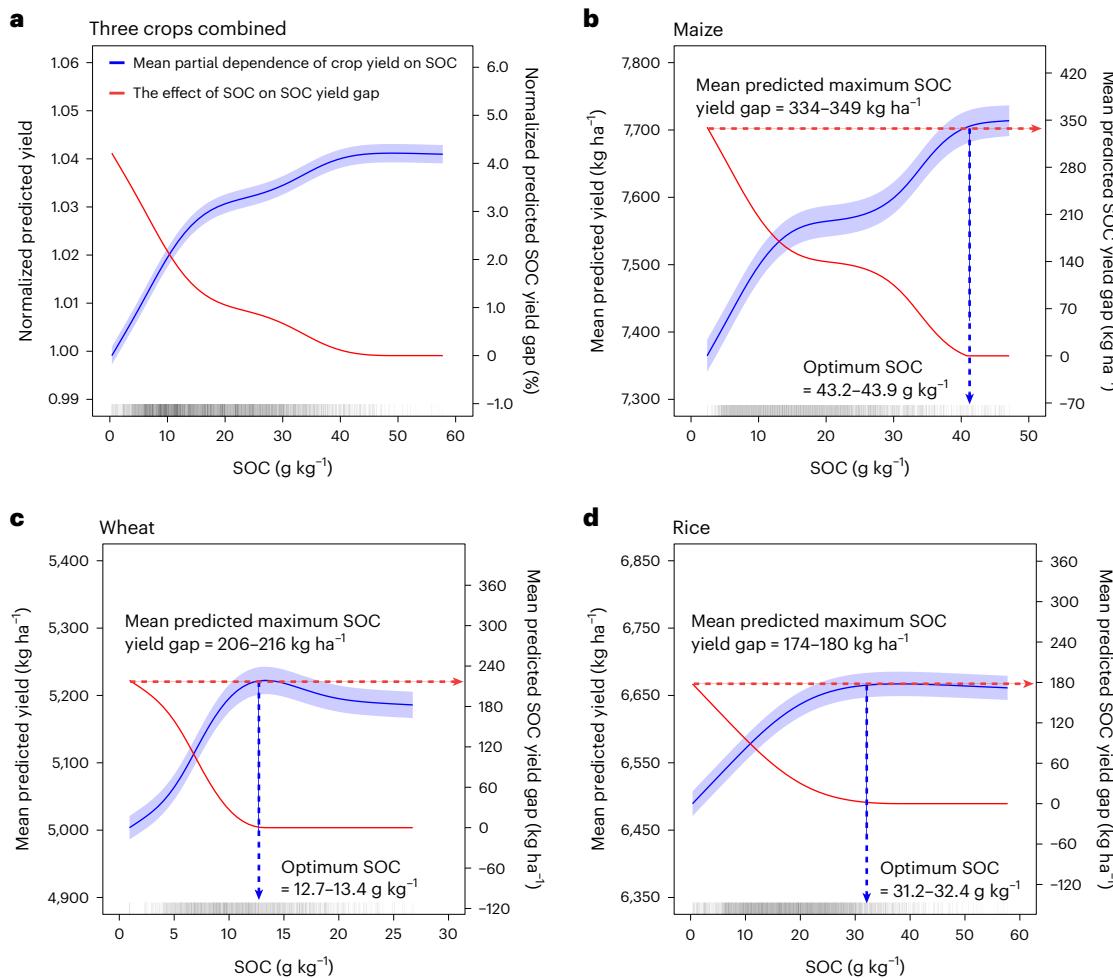


Fig. 1 | Causal effect of SOC on crop yield and SOC yield gap. **a**, Three crops combined. **b**, Maize. **c**, Wheat. **d**, Rice. The ribbons represent 95% confidence intervals. Optimum SOC is the SOC level beyond which the rate of yield increase is not different from zero ($p < 0.05$). The significance of the yield increase was estimated by two-sided t tests. The SOC yield gap is defined as the amount by

which yield could be increased by increasing SOC. For three crops combined, yield was normalized as the ratio of on-farm measured yield to predicted minimum yield for each crop, and SOC yield gap was normalized relative to minimum yield, expressed as percentage change. The grey bars at the bottom of each panel represent the distribution of SOC.

has been difficult to quantify for several reasons. First, there are many confounding factors that affect both yield and SOC that have been difficult to fully control or account for. These include, for example, climate, fertilization, land management, hydrology and soil properties. Second, the causal relationship between yield and SOC operates in both directions, making the causal impact of SOC on yield difficult to differentiate. For example, in addition to the expectation that SOC may increase yields, the opposite is also true: that higher yields correlate with greater organic matter inputs to the soil (from crop residues and belowground production), which in turn leads to higher SOC. Third, many surveys omit metadata on important determinants of yield, particularly in terms of farm management techniques such as tillage, weed control, pest control, irrigation, varietal selection and so on, which contribute to the large unexplained variance and a low signal-to-noise ratio in SOC yield relationships⁸. These three types of barrier have led to up to a half-century of ongoing debate about the quantitative importance of SOC in determining crop yields⁸. The resultant difficulty in quantitatively evaluating the SOC effect on yield has been a limiting factor in understanding the potential importance of SOC sequestration to global crop productivity, which therefore introduces large uncertainty into cost–benefit estimations and discourages risk-averse policymakers and farmers from adopting sustainable agricultural practices to sequester SOC⁷.

We addressed the above issues by analysing data from a programme of tens of thousands of controlled field trials with maize (number of field sites, $N_{\text{site}} = 4,906$), wheat ($N_{\text{site}} = 4,448$) and rice ($N_{\text{site}} = 4,308$), distributed across all major cropping regions of China covering a broad range of climate zones and soil types (Extended Data Fig. 1). Yield variation due to differences in management between farms was minimized by conducting recommended best management practices based on consistent on-site guidance and recommendations from local agricultural experts and trained extension personnel for key operations including weed, pest and disease control, cultivar, tillage, sowing date, seeding density, supplementary irrigation and harvest (Methods; description of field trials). Each field trial site had treatments with 0, 0.5, 1.0 and 1.5 times the optimum nitrogen (N) applications, combined with optimum phosphorus (P_2O_5) and potassium (K_2O) and a treatment with no fertilization. Climatic, edaphic, management and topographic covariates of SOC and yield were provided as explanatory variables for machine learning (both random forest and causal forest) models, thereby accounting for the most important confounding variables. Thus, the partial dependence of yield on SOC was quantified independently of such confounding variables and with variability in farm management controlled for (Methods; causal effect of SOC on crop yield). Field trials were almost ubiquitously conducted over a single season by researchers at on-farm locations, ensuring that observed

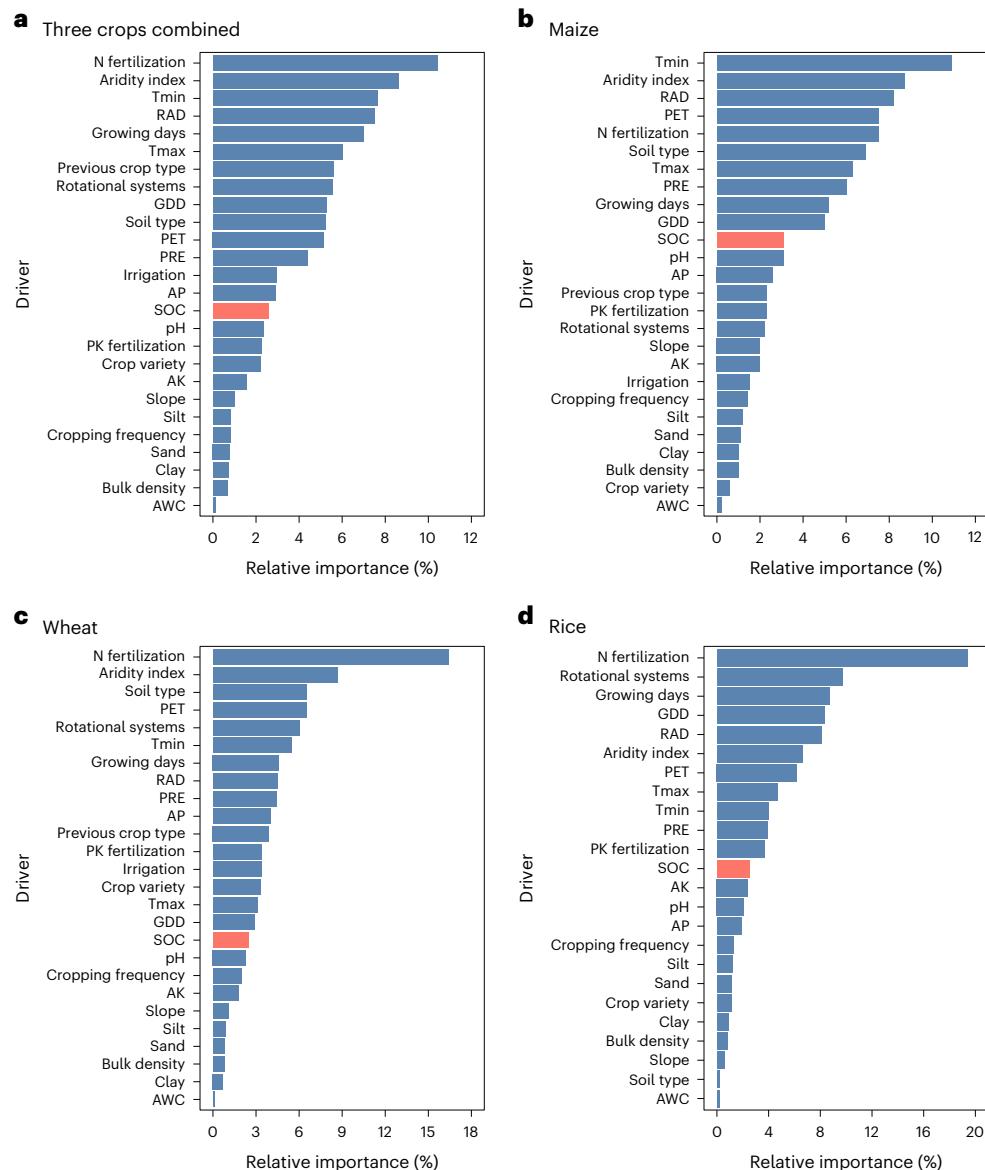


Fig. 2 | Drivers of crop yield. **a**, Three crops combined. **b**, Maize. **c**, Wheat. **d**, Rice. Bar charts indicate the relative importance (by permutation method; Methods) of variables as predictors of the crop yield. NPK fertilization = nitrogen, phosphorus and potassium fertilizer rate; PET = potential evapotranspiration;

GDD, RAD, PRE, Tmax and Tmin = growing degree days, radiation, precipitation, maximum mean temperature and minimum mean temperature during the growing days; AP = plant-available phosphorus; AK = plant-available potassium; AWC = available water capacity; SOC is highlighted in red for ease of reading.

variations in yield do not affect SOC whereas SOC can affect yield, thus providing confidence that the partial dependence of crop yield on SOC represents a causal effect that is distinct from a spurious correlation.

The causal effects of SOC and fertilizer on yield

On average, crop yields increased with SOC and levelled off as SOC approaches a threshold level, confirming the existence of an optimum SOC concentration beyond which little to no additional yield gains accrued from further increases in SOC^{8,11} (Fig. 1). We define the SOC yield gap as the difference between yield at the optimum SOC and yield at current SOC (SOC yield gaps were defined as zero when current SOC is above the optimum). The optimum SOC (in grams of organic carbon per kilogram of soil) was estimated as the SOC concentration beyond which the rate of increase in yield is not different from zero ($p < 0.05$). The mean optimum SOC was different for the three crops, namely 43.2–43.9 g kg⁻¹ (95% confidence interval (C.I.)) for maize, 12.7–13.4 g kg⁻¹ for wheat and 31.2–32.4 g kg⁻¹ for rice (Fig. 1). These

results expand previous studies that suggested 20 g kg⁻¹ as a universal average optimum SOC^{8,12}, underscoring the need to adopt specific optimum SOC levels for each cropping system, which also varies with soil and climate (Extended Data Fig. 2). Effect sizes estimated by the random forest model were in the same range as those estimated by causal forest, supporting that the random forest-based results are unbiased by heterogeneous treatment effects (Supplementary Text 1 and Table 1).

Nitrogen fertilizer input was, not surprisingly, the best predictor of yield for all three studied crops combined. Importantly, however, we found that SOC contributes to yield, with a relative prediction importance of around 3%, which is, on average, one-fifth that of nitrogen fertilization. The relative importance of nitrogen fertilization as a predictor of maize yield was half (8%) compared to wheat (16%) and rice (19%) (Fig. 2). This could be attributed to the lower nitrogen fertilizer use efficiency of maize compared with wheat and rice¹³. Previous studies have suggested that mineral fertilizers compensate to sustain yield even at low SOC levels¹⁴. We found that even though addition of more nitrogen

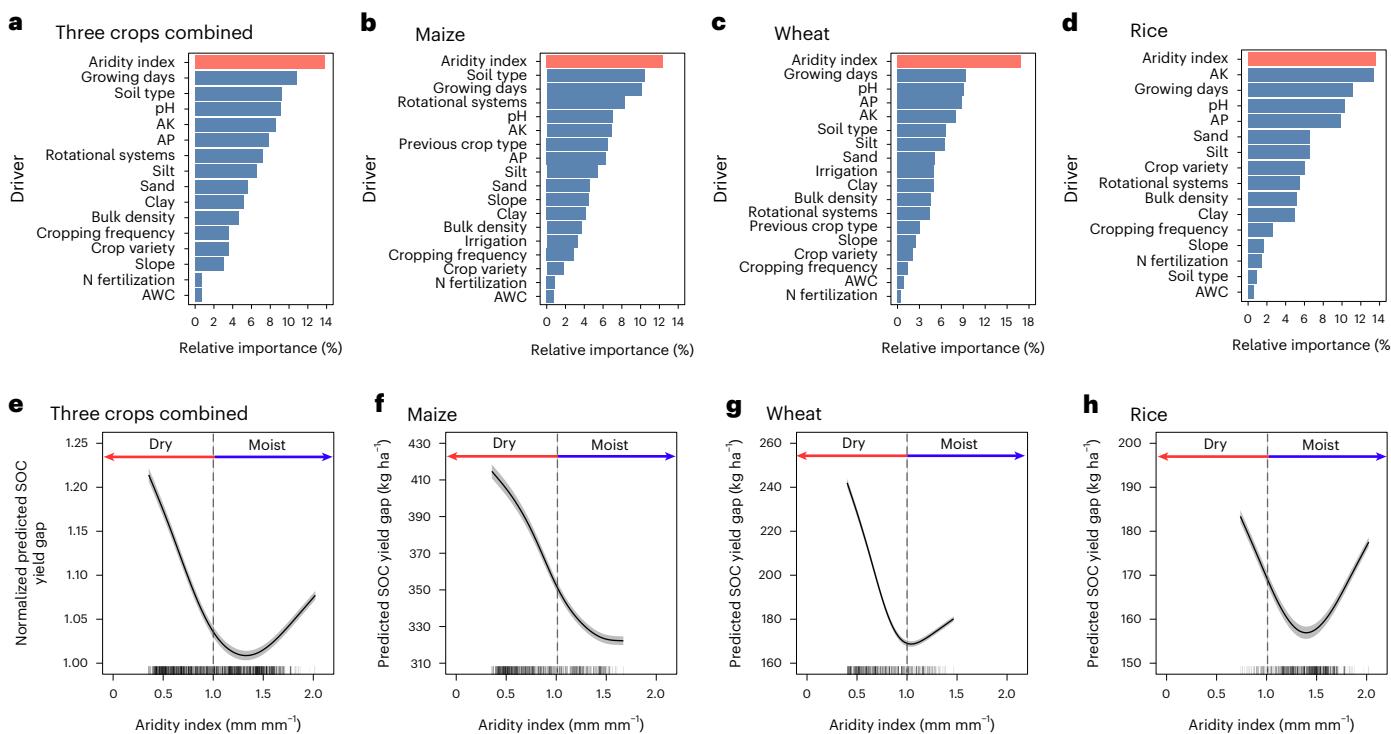


Fig. 3 | Drivers of SOC yield gap. **a–d**, Bar charts indicate relative importance of variables as predictors of SOC yield gap for three crops combined (**a**), maize (**b**), wheat (**c**) and rice (**d**). N fertilization = nitrogen fertilizer rate; AP = plant-available phosphorus, AK = plant-available potassium. AWC = available water capacity. **e–h**, Solid lines represent the mean partial dependence of the predicted SOC yield gap on the aridity index for three crops combined (**e**), maize (**f**), wheat (**g**) and rice (**h**) (lower aridity index values correspond to a drier climate). For

three crops combined, the predicted SOC yield gap was normalized relative to the predicted minimum SOC yield gap for each crop. The ribbons represent the 95% confidence interval. The dashed lines indicate the aridity index = 1.0 line, which differentiates dry versus moist climates; aridity index in panels **a–d** are highlighted in red for ease of identification; the grey bars at the bottom of panels **e–h** represent the distribution of aridity index data.

fertilizer (up to optimum rates) always contributed to higher crop productivity, crop productivity simultaneously increased with SOC up to the SOC optima under any nitrogen input rate (Extended Data Fig. 3 and Supplementary Text 2). This suggests that SOC increases may serve as a complementary strategy to nitrogen fertilizer amendments.

Feature importance for prediction of the SOC yield gap

The predicted maximum SOC yield gap (that is, difference between yields at the optimum SOC and the lowest SOC per climate zone) was generally higher in dry compared to moist climates. With maize, for example, the yield gap was fourfold greater in warm-temperate-dry compared to warm-temperate-moist climates (Extended Data Fig. 4 and Supplementary Text 3). This underlines the important role of aridity in controlling SOC yield gaps. Indeed, using a random forest model trained to predict SOC yield gaps rather than yield (Methods; drivers of soil organic carbon yield gap), aridity index was the most important parameter in determining this yield gap, explaining 12% of the variability in the SOC-derived yield gap for maize, 17% for wheat, 13% for rice and 14% for the three crops combined (Fig. 3). Most variation of the yield gap with aridity was found in dry climates (defined as an aridity index lower than unity) where this SOC-derived yield gap increased as soils became drier (Fig. 3). This indicates that improved plant-available water capacity (AWC) with greater SOC¹⁵ is the most important mechanism underlying observed increases in crop yield with higher SOC in the current large dataset. Indeed, SOC is known to contribute to the formation of aggregation and greater volume of pore spaces¹⁶, providing more space to store water from rainfall or irrigation to sustain crop yields during drought¹⁷. This result expands previous evidence of only a small increase in AWC for plants from SOC increase¹⁸,

as our results demonstrate that changes in AWC in response to a SOC increase can provide more moisture for crops to sustain 0.4–1.5% higher yield (16–45% of the SOC yield gap) under water-limited conditions.

Although most of the SOC yield gap response to aridity occurs in dry climates, a smaller increase was also observed (especially in rice) in moister and wetter climates, where water availability is less limiting to crop production (Fig. 3). This response observed for rice in moist climates could be explained by higher SOC decreasing cation leaching in wet regions via improved cation exchange capacity, which might therefore increase crop yield¹⁰.

Global and regional contribution of SOC to crop production

Crop yield models were obtained by random forest regression and used to estimate crop production increases across global maize-, wheat- and rice-producing regions for two scenarios. The first scenario corresponds to an unconstrained increase in SOC up to its yield-maximizing optimum (Methods; global extrapolation of SOC yield gap). Note that this optimum level varies with crop, climate and soil (Extended Data Fig. 2). Although there is no known pathway to achieving this optimum SOC in most locations, the unconstrained scenario allows us to bracket the maximum theoretical yield impact from SOC should future advances allow SOC to be increased far beyond the current technical potential. Such advances could include, for example, enhanced phenotypes¹⁹ or higher photosynthetic efficiency²⁰. The unconstrained scenario also allows us to locate the greatest impacts of improved SOC, recognizing that although achieving this unconstrained potential everywhere simultaneously may not be feasible, understanding where the greatest positive effects from organic matter additions to soil can be realized may assist in optimizing spatial application of transportable organic

Table 1 | Global production increase and calories provided from closing SOC yield gaps for three studied crops

	Global production increase for maize (Tg yr ⁻¹)	Global production increase for wheat (Tg yr ⁻¹)	Global production increase for rice (Tg yr ⁻¹)	Total production increase for the three crops (Tg yr ⁻¹)	Energy provision from production increase for the three crops (kcal yr ⁻¹)
Technical potential	10	6	4	20	71×10^{12}
Unconstrained	50	58	12	120	420×10^{12}

Technical potential represents yield gains when SOC increased to levels that are achieved using a combination of high organic matter inputs and zero tillage; unconstrained increase in SOC represents yield gains when SOC increased to optimum levels.

resources. The second scenario corresponds to the technical potential by which SOC stocks could be increased from current levels using currently available management options including a combination of high organic matter inputs (from cover crops and crop residue retention) and zero tillage²¹ (Methods; global extrapolation of SOC yield gap).

Raising SOC to optimum levels (the unconstrained scenario), were it achievable, would increase global maize, wheat and rice potential production (once other management and fertilization constraints have already been eliminated) by 120 Tg yr⁻¹ in aggregate (Table 1), with the global crop production increasing by 4.3%. This is equivalent to 700 million people's cereal demand and the energy needs for 640 million people (420×10^{12} kcal yr⁻¹). The unconstrained SOC yield gap was greatest in dry climates, being on average 1.3–2.2 times greater than in moist and wet climate zones (Supplementary Table 2). Specifically, the US maize belt; Central, South and Southeast Asia; eastern Asia; maize-producing regions in Canada; South America; Europe; sub-Saharan Africa and Russia have the largest quartile of SOC-derived maize yield gaps, with 549–1,016 kg ha⁻¹. For wheat, the 25% largest SOC-derived yield gaps are observed in the southwestern United States; northern Mexico; North Mediterranean basin; North China Plain; wheat-producing regions in South America; sub-Saharan Africa; Central, South and Southeast Asia and Australia, with 407–898 kg ha⁻¹. Rice-producing regions demonstrate only a slight increase in yield when increasing SOC in an unconstrained way, with the 25% largest yield increase of 192–291 kg ha⁻¹ found in southern and eastern South America; Central, West and South Asia; eastern Asia; rice-producing regions in sub-Saharan Africa and Australia (detailed regional distribution of the 25% highest SOC yield gaps is listed in Fig. 4, Extended Data Fig. 5 and Supplementary Table 3, respectively). The low SOC yield gaps for rice can be explained by the current SOC levels in flooded rice systems already approaching yield-optimizing levels.

The technical potential for SOC sequestration is, however, substantially smaller than the unconstrained scenario, with potential global maize, wheat and rice production increasing by only 0.7% (20 Tg yr⁻¹), once other management and fertilization constraints have also been addressed (Table 1). The average weighted global SOC yield gap for the technical potential is 41 (± 4 ; one standard error, estimated by the jackknife method in Extended Data Fig. 6) kg ha⁻¹ for all three crops (Methods; uncertainty analysis of global extrapolations). However, this SOC yield gap is somewhat higher for specific crops in specific regions (Fig. 4), with the 25% highest SOC yield gaps ranging from 73 ± 2 kg ha⁻¹ for rice in Central, West and Southeast Asia to 231 ± 4 kg ha⁻¹ for maize-producing regions in sub-Saharan Africa (detailed regional distribution with the 25% highest SOC yield gaps is listed in Extended Data Fig. 5 and Supplementary Table 3). Closing yield gaps by improving SOC to the technical potential, once other constraints have been removed, could contribute to an annual increase of 71×10^{12} kcal energy for the three crops globally, which would meet 120 million people's cereal demand, or caloric needs of 110 million people.

It is important to note that the global extrapolation assumes optimum management practices where the yield constraints derived from inadequate nutrition, hydration, pest, weed and disease have already been addressed. Nonetheless, addressing management and soil quality

constraints are complementary approaches, whereby increasing soil quality without improving management or vice versa is expected to give lower absolute yield improvements²². Furthermore, although the field trial data were collected from a diverse geographic extent that includes seven of eight climate zones, ten of the 12 soil texture classes used by the US Department of Agriculture (USDA) and three of the four major soil orders (Alfisols, Inceptisols, Mollisols and Ultisols) where global maize-, wheat- and rice-producing regions are located, a few of the climate and soil regions where these crops are produced had only a few or no field trial data, especially in tropical dry and tropical montane climate zones and Alfisol regions. Extrapolations to these underrepresented regions should therefore be treated with more caution than those regions with more data. Nonetheless, because the field trials cover the most important climate zones for these crops globally (Extended Data Fig. 7) and soils with 10 out of 12 USDA soil texture classes, pH values of 3–9, SOC of 1–58 g kg⁻¹ exceeding the average range observed for global producing regions of these crops, the overall conclusions are well supported.

Implications for global soil organic carbon sequestration

Despite the expectation expressed in earlier publications that SOC can make large (up to 10–23%) contributions to crop production increases^{8,23} and that crop yield increases would be therefore be a driving force behind the adoption of SOC sequestration at a global scale^{6,7}, our finding that the technical potential provides less than 1% increase in global crop production challenges this assumption and rather suggests adopting a more nuanced regional focus. The negligible increase in global average crop production is probably lower than previous estimates both as a result of considering currently available technologies to improve SOC and also because we have isolated the causal effect from an overall correlation. Although crop yields under the technical potential increase only by a global average of 41 kg ha⁻¹, within countries that have a weak or moderate food security, the mean impact is higher (for example, 71 kg ha⁻¹ in sub-Saharan Africa, 106 kg ha⁻¹ in Central America, 89 kg ha⁻¹ in the Caribbean and 72 kg ha⁻¹ in South America) (Extended Data Fig. 8c and Supplementary Text 4). Notably, yield improvements derived from SOC, once other constraints have been addressed, vary from 26% (technical potential) to 86% (unconstrained potential) of current food deficits in food-insecure countries (Extended Data Fig. 9 and Supplementary Text 4). Because the SOC impacts on yield reported here are derived from a dataset where optimum management and fertilization were already implemented, the extent to which food deficits could be alleviated by SOC under current situations in food-insecure countries remains an open question. Importantly, SOC benefits accrue gradually over decades. Therefore, adoption of integrated nitrogen fertilizer, pest, weed and SOC management strategies may be the most promising means to improve crop production in food-insecure countries because their benefits are both more immediate and larger.

Although unconstrained increases in SOC up to their yield-maximizing level would provide a much more compelling yield benefit, being six times larger than the technical potential, such large increases in SOC are not achievable using agricultural practices that can currently be applied at scale. This remaining SOC yield

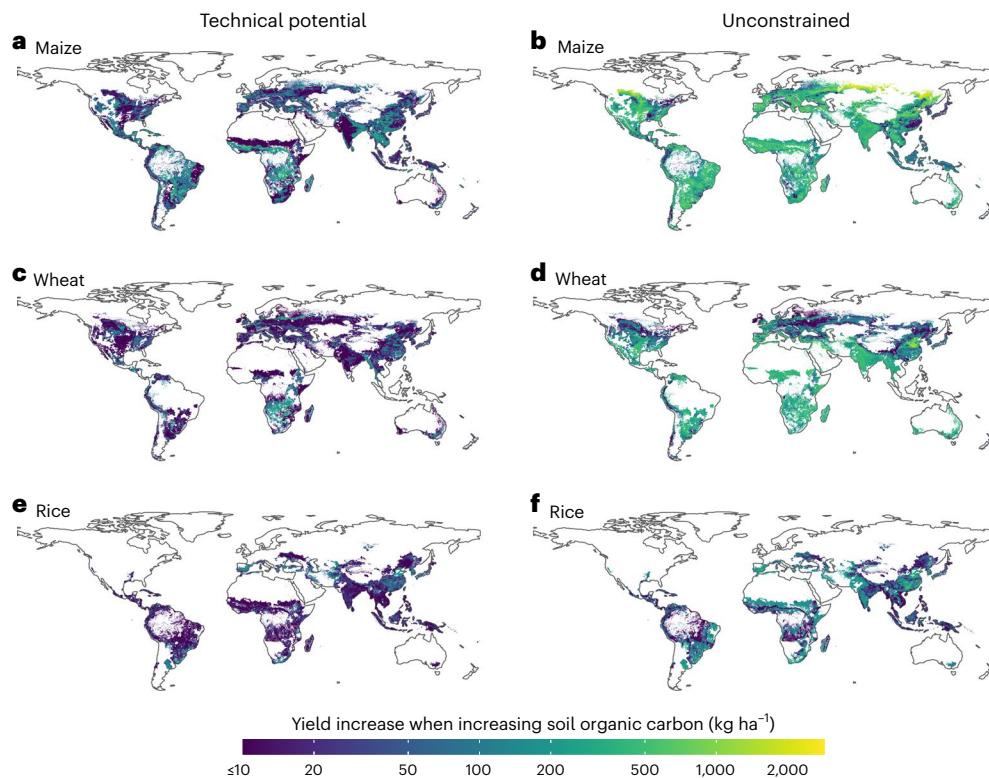


Fig. 4 | Global potential to increase yields by increasing SOC. **a,c,e**, Technical potential indicates the amount by which crop yields could be increased by increasing SOC to its current technical potential SOC levels for maize (**a**), wheat (**e**) and rice (**e**). Current technical potential SOC levels indicate the levels of SOC that can be achieved using a combination of high organic matter inputs

(including cover crops and crop residues) and zero tillage. **b,d,f**, Unconstrained indicated the amount by which crop yields could be increased by increasing SOC to its soil- and climate-specific optimum levels for maize (**b**), wheat (**d**) and rice (**f**) (Extended Data Fig. 2). Basemap reproduced from ArcGIS/Esri.

gap, beyond what can be achieved with current scalable practices, provides a strong rationale to investigate how much further we can extend the envelope of achievable SOC sequestration through use of novel practices beyond cover cropping, zero tillage and residue returns. Other possible options for sequestering greater amounts of organic carbon in soils include biochar²⁴; plant breeding for increased photosynthetic efficiency²⁵ and increased belowground production²⁰; or changing the soil microbiome to favour carbon-stabilizing taxa²⁶. Furthermore, the large geographic variation in SOC yield gaps with climate and soil types in the unconstrained scenario suggests potentially realizable benefits from redistributing available carbon resources to regions with the greatest yield increases or with large food security co-benefits. However, such a practice would need to be balanced against the possibility of soil degradation in areas from which organic matter is removed and the logistic cost of transporting large quantities of material.

Carbon dioxide removal practices, such as SOC sequestration, must be implemented and scaled up rapidly if we are to reach net zero GHG emissions by mid-century. According to our analysis, improving soil fertility using currently available technologies is unlikely to be a major driver for adoption of SOC sequestration in much of the world, and other policy measures would need to be explored if the full potential for SOC to address climate change is to be achieved. Nonetheless, the motivation of enhancing crop production in certain regions or the promise of SOC to help address food deficits in food-insecure countries could prove more likely to drive SOC sequestration policy in some regions. Furthermore, novel practices that extend the boundary of achievable SOC sequestration to optimum levels are important areas of further research that could drive climate policy by bringing larger yield benefits within reach.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41561-023-01302-3>.

References

- IPCC Climate Change 2021: The Physical Science Basis (eds Masson-Delmotte, V. et al.) (Cambridge Univ. Press, 2021).
- IPCC Special Report on Global Warming of 1.5 °C (eds Masson-Delmotte, V. et al.) (WMO, 2018).
- Anderson, C. M. et al. Natural climate solutions are not enough. *Science* **363**, 933–934 (2019).
- Field, C. B. & Mach, K. J. Rightsizing carbon dioxide removal. *Science* **356**, 706–707 (2017).
- Batjes, N. H. Total carbon and nitrogen in the soils of the world. *Eur. J. Soil Sci.* **47**, 151–163 (1996).
- Minasny, B. et al. Soil carbon 4 per mille. *Geoderma* **292**, 59–86 (2017).
- Amelung, W. et al. Towards a global-scale soil climate mitigation strategy. *Nat. Commun.* **11**, 5427 (2020).
- Oldfield, E. E., Bradford, M. A. & Wood, S. A. Global meta-analysis of the relationship between soil organic matter and crop yields. *Soil* **5**, 15–32 (2019).
- Lal, R. Soil carbon sequestration impacts on global climate change and food security. *Science* **304**, 1623–1627 (2004).
- Lal, R. Soil organic matter content and crop yield. *J. Soil Water Conserv.* **75**, 27A–32A (2020).

11. Lal, R. Enhancing crop yields in the developing countries through restoration of the soil organic carbon pool in agricultural lands. *Land Degrad. Dev.* **17**, 197–209 (2006).
12. Pan, G., Smith, P. & Pan, W. The role of soil organic matter in maintaining the productivity and yield stability of cereals in China. *Agric. Ecosyst. Environ.* **129**, 344–348 (2009).
13. Yu, X. et al. Global meta-analysis of nitrogen fertilizer use efficiency in rice, wheat and maize. *Agric. Ecosyst. Environ.* **338**, 108089 (2022).
14. Oelofse, M. et al. Do soil organic carbon levels affect potential yields and nitrogen use efficiency? An analysis of winter wheat and spring barley field trials. *Eur. J. Agron.* **66**, 62–73 (2015).
15. Hudson, B. D. Soil organic matter and available water capacity. *J. Soil Water Conserv.* **49**, 189–194 (1994).
16. Murphy, B. Impact of soil organic matter on soil properties—a review with emphasis on Australian soils. *Soil Res.* **53**, 605–635 (2015).
17. Kane, D. A., Bradford, M. A., Fuller, E., Oldfield, E. E. & Wood, S. A. Soil organic matter protects US maize yields and lowers crop insurance payouts under drought. *Environ. Res. Lett.* **16**, 044018 (2021).
18. Minasny, B. & McBratney, A. Limited effect of organic matter on soil available water capacity. *Eur. J. Soil Sci.* **69**, 39–47 (2018).
19. Paustian, K. et al. Climate-smart soils. *Nature* **532**, 49–57 (2016).
20. DeLisi, C. et al. The role of synthetic biology in atmospheric greenhouse gas reduction: prospects and challenges. *Biosci. Res.* **2020**, 1016207 (2020).
21. Ogle, S. M. et al. in *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Agriculture, Forestry and Other Land Use*, Vol. 4 (eds Calvo Buendia, E. et al.) Ch. 2 (IPCC, 2019).
22. Kimetu, J. M. et al. Reversibility of soil productivity decline with organic matter of differing quality along a degradation gradient. *Ecosyst* **11**, 726–739 (2007).
23. Oldfield, E. E. et al. Positive associations of soil organic matter and crop yields across a regional network of working farms. *Soil Sci. Soc. Am. J.* **86**, 384–397 (2022).
24. Lehmann, J. et al. Biochar in climate change mitigation. *Nat. Geosci.* **14**, 883–892 (2021).
25. South, P. F., Cavanagh, A. P., Lopez-Calcagno, P. E., Raines, C. A. & Ort, D. R. Optimizing photorespiration for improved crop productivity. *J. Integr. Plant Biol.* **60**, 1217–1230 (2018).
26. Kroeger, M. E. et al. Microbial community composition controls carbon flux across litter types in early phase of litter decomposition. *Environ. Microbiol.* **23**, 6676–6693 (2021).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

Methods

Description of field trials

Field trials were conducted from 2005 to 2013 during the National Soil Test and Fertilizer Recommendation Projects, which were previously described in ref. 27. In the current study, field trials from a total of 13,662 sites (number of field sites are 4,906 for maize, 4,448 for wheat and 4,308 for rice) were collected (Extended Data Fig. 1), and the dataset was established by expanding multiple biophysical factors and management measurements to quantify the causal effect of SOC on crop yield. All trials were designed and managed by local agricultural experts and/or trained extension personnel and were implemented in on-farm fields based on a stratified randomized sample of local farms according to a series of criteria including selection of local typical soil types, a broad range of soil fertility gradients and fertilization levels and crop type. Nearly all field trials were conducted over a single season (less than 3% of the trials were conducted for two or three seasons). Five nitrogen fertilizer application rates were used in each trial, including 0, 0.5, 1.0 and 1.5 times the optimum nitrogen rate at optimum phosphorus and potassium rates; and a zero-fertilization treatment without nitrogen, phosphorus or potassium additions (control). Optimum nitrogen, phosphorus and potassium rates were recommended by local agricultural experts and/or trained extension personnel based on an integrated nutrient management strategy²⁸, with the target to maximize both nutrient-use efficiency and crop yield for a given site. No organic amendments or straw were applied during the trials. In total, there were 66,593 treatments (number of treatments are 24,107 for maize, 20,932 for rice and 21,554 for wheat) recorded during the project when missing treatments were removed (1,717 treatments were removed due to missing yield data because of recording issues or bad weather). Local agricultural experts and/or trained extension officers provided direct and consistent on-site guidance and recommended management advice during key farming operations, including weed, pest and disease control; cultivar selection; tillage; sowing date and density; irrigation (only in irrigated systems, and all rice trials were flood-irrigated) and harvest, to minimize the impact of management variability other than fertilization on variation in crop yield. Thus, the field trial design minimized both management constraints to yield and variability in yield due to management. Prior rotational systems, cropping frequency and previous crop type were recorded by local agricultural experts and/or trained extension officers through interviews with farmers before the trial. Previous crop type was not recorded for rice.

Metadata characterizing the site and management during the field trials were collected by the local experts conducting the trials. The information recorded included irrigation, crop variety, slope, planting and harvest dates. Yield was measured at crop physiological maturity and was standardized to an assumed 14% moisture content for maize and rice and 13% for wheat. Climatic data during crop growing periods, including daily average maximum (Tmax) and minimum (Tmin) temperatures, precipitation (PRE) and sunshine duration (SSD), were collected for each field trial by using the daily climatic data from the Chinese Meteorological Administration in the county or municipality where the trial was implemented. Daily Solar Radiation (RAD) was calculated based on SSD using the WeatherAid module in the Hybrid-Maize model²⁹. Growing degree days (GDD) were calculated as the annual sum of days in which the daily mean temperature exceeded 10 °C for maize and rice and 0 °C for wheat. Aridity index was calculated as mean annual precipitation divided by mean annual potential evapotranspiration (PET) using data for the year of the trial collected from the Climate Research Unit (CRU) Time-Series version 4.04 of high-resolution gridded data of month-by-month variation in climate³⁰.

Soil sampling for analysis of SOC and other soil indicators

Soil samples were collected from a depth of 0–0.3 m. Five soil samples were collected from each field trial and then combined for laboratory measurements of soil properties, including soil organic carbon (SOC),

soil pH, plant-available potassium (AK) and plant-available phosphorus (AP), based on standard methods³¹. Specifically, SOC was measured by the potassium dichromate volumetric method. Soil pH was measured with a glass electrode in a 1:2.5 soil/water suspension. AK was analysed by a flame photometer after extraction with 1 M ammonium acetate. AP was analysed by the Olsen method. Soil type was classified according to the Chinese soil taxonomy system³². Soil bulk density and relative proportion of clay, silt and sand in the topsoil for each field trial were estimated geospatially using SoilGrids version 2.0³³. Available water capacity (AWC) was estimated with texture pedotransfer functions³⁴ in the topsoil.

Causal effect of SOC on crop yield

Random forest regression models were established for each of the three crops (maize, wheat and rice) separately, using the R package ‘ranger’³⁵, with the above climatic, edaphic, topographic, fertilization and management variables as independent predictors. Additionally, a combined random forest model for normalized yield of all three crops was derived using the same independent variables, with normalized yield calculated by dividing in-field measured crop yield by predicted minimum yield of the crop type (that is, when SOC is at the low end of the range for each crop) in Fig. 1 (7,365 kg ha⁻¹ for maize, 5,033 kg ha⁻¹ for wheat, 6,472 kg ha⁻¹ for rice). The independent variables for fertilization rates were expressed relative to optimum rates, with values of 0, 0.5, 1 and 1.5 for N and values of 0 and 1 for P and K. Parameter importance was calculated using the permutation method³⁶. We set the ‘respect.unordered.factors’ parameter in ranger equal to ‘TRUE’ to ensure that unordered categorical data are not assumed to be ordered relative to regression results. Coefficients of determination (R^2 , estimated using the internal out-of-bag samples) and index of agreement (d) were used to assess model performance. The performance of the yield models (both R^2 and d) refers to part 1 in Supplementary Table 4.

Partial dependence of the random forest was adopted to eliminate confounding factors, which established the mean causal effect of SOC on crop yield for each separate studied crop and the three crops combined. Partial dependence was calculated using the R package ‘pdp’³⁷ and smoothed using Tikhonov regularization (ridge regression)³⁸. Leave-one-out cross validation was adopted to adjust the regularization parameter used for smoothing the partial dependence curves. Partial dependence variance and standard error were calculated using bootstrap sampling of model predictions. The optimum SOC was determined by the concentration beyond which the rate of increase in yield with SOC is not different from 0 ($p < 0.05$). The significance of yield increases was estimated by two-sided t test. The 95% C.I. of the optimum SOC was determined by bootstrapping. A causal forest regression model was used to validate the partial dependence of the random forest method by comparing the average treatment effect size of SOC on crop yield for the two methods (Supplementary Text 1 provides results of the causal forest validation). The causal forest was derived using the R package ‘grf’³⁹. Furthermore, scatter plots of the measured data for SOC and crop yield of the three crops are provided in Supplementary Fig. 1.

Partial dependence of the random forest was also used to investigate the causal effect of SOC on crop yield for each crop under different nitrogen fertilizer rates based on above steps and above climate, soil and management variables. For the three crops combined under different nitrogen fertilization rates (Extended Data Fig. 3), predicted yield was normalized to predicted minimum yield for each crop under the same treatment.

Causal effect of SOC on crop yield under climate zones

Field trials were grouped based on the Intergovernmental Panel on Climate Change (IPCC) climate zone definitions⁴⁰ classified according to elevation and climatic data³⁰. Partial dependence of the random forest was used to investigate the causal effect of SOC on crop yield

for each crop under different climate zones based on above steps and above climate, soil and management variables. For each climate zone (Extended Data Fig. 4), normalized predicted yield was calculated as the ratio of predicted yield from partial dependence to predicted minimum mean yield for each crop under each climate zone, which facilitated comparisons between climate zones.

Drivers for the SOC yield gap

The above maize, wheat and rice yield random forest models were used to estimate the SOC yield gaps disaggregated into specific crop, soil and climate combinations, rather than the mean SOC yield gap across all locations derived from Fig. 1. In each field trial, the SOC yield gap was defined as the difference between predicted maximum yield at optimum SOC level and predicted yield at the current SOC level. The climate, soil and management independent variables in the dataset were provided as inputs to the random forest yield models to estimate crop yield at current SOC for each trial. The predicted maximum yield at optimum SOC was calculated for each field trial as the maximum yield when increasing SOC while holding other independent variables constant.

Random forest yield models were trained to predict the SOC yield gap for each trial of maize, wheat and rice systems separately and the normalized yield gap for all crops combined, using R ranger as described above for the yield models. The normalized SOC yield gap of all three crops in the combined random forest model was calculated by dividing estimated SOC yield gap in each of the field trials by the predicted minimum SOC yield gap of the crop type (that is, when aridity index is at the low end of the range for each crop) from Fig. 3 (322 kg ha⁻¹ for maize, 169 kg ha⁻¹ for wheat, 157 kg ha⁻¹ for rice). Coefficients of determination (R^2 , estimated using internal out-of-bag samples) and index of agreement (d) were used to assess SOC yield gap model performance (part 2 in Supplementary Table 4). Partial dependence of the SOC yield gap for which the highest-importance predictor (aridity index) was calculated as described for partial dependence in the yield model above. For partial dependence of the SOC yield gap on aridity index, parameters that showed a strong correlation with aridity index (that is, GDD, RAD, PET, Tmax, Tmin and PRE) were removed from the models, because varying aridity index while holding these strongly correlated variables constant would lead to inaccurate predictions. Effects of P and K fertilization were also removed, due to their negligible importance (less than 0.1% explanation of the SOC yield gap). The maximum contribution of aridity index on the SOC yield gap (or the contribution of AWC to the SOC yield gap) was calculated for each crop as the difference between maximum and minimum SOC yield gap in the line charts of Fig. 3, which was then used to estimate the contribution to the absolute crop yield and total SOC yield gap. The scatter plot of the individual data for aridity index and SOC yield gap of the three crops is provided in Supplementary Fig. 2.

Global extrapolation of SOC yield gaps

SOC yield gaps were extrapolated to global maize-, wheat- and rice-producing regions. Gridded global data on cropped area for each of these crops were from ref. 41 (at 0.08° spatial resolution). Climate and soil data were resampled to this same target resolution using bilinear interpolation. Annual climate data (PRE, Tmax, Tmin, aridity index and PET) were from the CRU Time-Series version 4.04³⁰. The yield model for global extrapolation was trained using annual mean CRU TS4.04 climate data corresponding to the specific years in which each field was conducted. For global prediction, we used climate data averaged over the years 2011–2019. Soil properties (bulk density, clay, silt, sand, pH) were from SoilGrids v2.0³³. AWC was estimated with texture pedotransfer functions³⁴.

Both current SOC in croplands and its technical potential to be increased under currently available technologies using high organic matter inputs (residue retention and cover crops) and no till were estimated using the SOCTier 1 model of the 2019 revisions to the IPCC

guidelines for national GHG inventories²¹, according to the method of ref. 42.

These climate and soil independent variables were provided as inputs to the random forest yield models (the random forest yield models having been trained using field trial data) to predict the increase in yield of each crop as SOC is increased from current to either optimum or technical potential values. Optimum SOC (Extended Data Fig. 2) values were defined as the SOC level at which crop yield is maximized when increasing SOC in the global extrapolation section. Because the training data are from field trials that used best management practices to minimize variance from management, it was assumed that optimum fertilization and management would be used globally. Thus, the global extrapolation estimates the additional yield that could be achieved through higher SOC once yield gaps from suboptimal fertilization or management have already been addressed. Coefficients of determination (R^2 , estimated using the internal out-of-bag samples) and index of agreement (d) were used to assess model performance (part 3 in Supplementary Table 4). The yield gap was defined as zero if the yield at unconstrained SOC levels was less than the yield at current levels because only yield increases from increasing SOC were considered.

Global total SOC yield gaps were calculated as the sum of the yield gaps multiplied by the harvested area per crop in each pixel⁴¹. The contribution of yield increase for the three crops to the global crop production was estimated by using the yield increases under the two scenarios, divided by the global annual production of the sum of the three studied crops (2.8 Pg on a dry mass basis) in 2021⁴³. The additional per capita cereal supply was calculated assuming an annual cereal consumption of 171 kg per capita⁴⁴. The SOC yield gap was also converted to energy equivalents, assuming calorific values of 3.65 kcal g⁻¹ (maize), 3.39 kcal g⁻¹ (wheat) and 3.60 kcal g⁻¹ (rice)⁴⁵ on a moist mass basis. The number of additional people's energy needs that could be provided was then calculated assuming a mean daily energy requirement of at least 1,800 kcal per capita⁴⁶.

Crop production increases in each country were calculated as the total yield increase of maize, wheat and rice as a result of both technical potential and unconstrained increases in SOC. The country-level crop production increase from SOC increases were then expressed relative to annual current production of the three crops in each country to total harvested areas of the three crops in each country and to annual production of the three crops at optimum management. Annual current production of the three crops was calculated as the mean yield per hectare multiplied by the sum of country-level harvested areas of the three crops, where the unit yield was the mean of values from 2018 to 2020⁴³, and the country-level harvested areas were calculated as the sum of the harvested areas⁴¹ in a country.

Uncertainty analysis of global extrapolations

To investigate prediction uncertainty, standard errors from prediction for maize, wheat and rice yield models were estimated using the jackknife method (leave-one-out cross validation), with a random forest model trained separately on each data iteration. Each of these jackknife random forest models was then used to predict the yield gap with the parameter values of the omitted data point at each SOC value from zero to 50 g kg⁻¹ (in 0.1 g kg⁻¹ steps). Finally, mean predictions and their standard errors were calculated for each of the SOC values. The maximum standard errors for prediction of the SOC yield gap via SOC changes was 4 kg ha⁻¹ for maize, 3 kg ha⁻¹ for wheat and 2 kg ha⁻¹ for rice (Extended Data Fig. 6).

Benchmarking of yield potential against current food deficits

Annual total caloric food deficit in each country was calculated using the depth (size) of the food deficit (kcal per capita per day) multiplied by the number of the undernourished population and the number of days in a year. The energy content of in-country crop production increase of the three crops (under either the technical potential or the

unconstrained SOC increase) was compared to food deficits by dividing by the total per country food deficit.

The undernourished population in each country was calculated as the prevalence of undernourishment (2018–2020)⁴⁷ multiplied by the total population (2018–2020)⁴⁸. Prevalence of undernourishment or food deficit in each country was defined as the undernourished population as a proportion of the total population. The depth of daily food deficit (per capita per day) in each country was estimated as the difference between the daily minimum dietary energy requirement and the daily mean dietary energy consumption of the undernourished population⁴⁹. These are the most recent published data providing global per country values for depth of food deficit. Because these data provide the depth of food deficit per malnourished person, they are constrained by human physiology to lie in the range of up to a few hundred calories per capita per day below the minimum dietary energy requirement. Food deficits greater than this cannot be sustained for longer than short periods during famine, which is not considered in our model. Our estimates based on year 2000 data are, accordingly, expected to be conservative rather than over-estimates in most countries, as the general trend since then has been towards increased production. In contrast to the depth of food deficit, the prevalence of undernourishment changes greatly over time and is the most dominant factor that influences total food deficit and is the value for which the most recent data (2018 to 2020) are available⁴⁷.

For this benchmarking of SOC yield increases against food deficits, only in-country production and deficits were considered, rather than imports and exports. For countries where the additional caloric production exceeded the food deficit, the fraction of food deficit addressed was capped at 100%. Countries were classified using their food security level based on the Global Food Security Index⁵⁰, with five cohorts corresponding to ‘weak’, ‘moderate’, ‘good’ and ‘very good’ food security environments and a cohort of countries for which food security levels were not provided. The countries for which food security levels were not provided were then reclassified into above four classes based on the quartile according to prevalence of undernourishment (Supplementary Table 6). Finally, only the countries with ‘weak’ and ‘moderate’ food security levels were defined as food-insecure countries (for all cohorts in Supplementary Table 6), because the food deficits are not considered a major issue in these countries with ‘good’ and ‘very good’ food security levels (the prevalence of undernourishment in those countries are recorded as less than 2.5%).

Data availability

Data that support these findings are available via figshare (<https://doi.org/10.6084/m9.figshare.24137280>).

Code availability

Codes for processing the data are available via figshare (<https://doi.org/10.6084/m9.figshare.24137280>).

References

27. Qiao, L. et al. Soil quality both increases crop production and improves resilience to climate change. *Nat. Clim. Change* **12**, 574–580 (2022).
28. Zhang, F. et al. Integrated nutrient management for food security and environmental quality in China. *Adv. Agron.* **116**, 1–40 (2012).
29. Yang, H. S. et al. Hybrid-maize—a maize simulation model that combines two crop modeling approaches. *Field Crops Res.* **87**, 131–154 (2004).
30. Harris, I. C., Jones, P. D. & Osborn, T. CRU TS4.04: Climatic Research Unit (CRU) Time-Series (TS) Version 4.04 of High-Resolution Gridded Data of Month-by-Month Variation in Climate (Jan. 1901–Dec. 2019) (Univ. of East Anglia Climatic Research Unit, 2020); <https://catalogue.ceda.ac.uk/uuid/89e1e34ec3554dc98594a5732622bce9>
31. Jiang, R. F. & Cui, J. Y. in *Fertilization Technology Highlights* (Ed. Zhang, F. S.) Ch. 5 (China Agricultural Univ. Press, 2006).
32. National Soil Survey Office *Chinese Soil* (China Agricultural Press, 1998).
33. Poggio, L. et al. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil* **7**, 217–240 (2021).
34. Dobarco, M. R., Cousin, I., Le Bas, C. & Martin, M. P. Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. *Geoderma* **336**, 81–95 (2019).
35. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
36. Altmann, A., Tološi, L., Sander, O. & Lengauer, T. Permutation importance: a corrected feature importance measure. *Bioinform* **26**, 1340–1347 (2010).
37. Greenwell, B. M. pdp: an R package for constructing partial dependence plots. *R J.* **9**, 421–436 (2017).
38. Eilers, P. H. A perfect smoother. *Anal. Chem.* **75**, 3631–3636 (2003).
39. Altey, S. & Wager, S. Estimating treatment effects with causal forests: an application. *Obs. Stud.* **5**, 37–51 (2019).
40. Reddy, S. et al. in *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Agriculture, Forestry and Other Land Use*, Vol. 4 (eds Calvo Buendia, E. et al.) Ch. 3 (IPCC, 2019).
41. Monfreda, C., Ramankutty, N. & Foley, J. A. Farming the planet: 2. geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Glob. Biogeochem. Cycles* **22**, GB1022 (2008).
42. Roe, S. et al. Land-based measures to mitigate climate change: potential and feasibility by country. *Glob. Change Biol.* **27**, 6025–6058 (2021).
43. FAOSTAT: *Crops and Livestock Products* (FAO, 2021); <https://www.fao.org/faostat/en/#data/QCL>
44. Alexandratos, N. & Mernies, J. in *World Agriculture: Towards 2015/2030 An FAO Perspective: Prospects for Food and Nutrition* (Ed. Bruinsma, J.) Ch. 2 (FAO, 2003).
45. Nuss, E. T. & Tanumihardjo, S. A. Maize: a paramount staple crop in the context of global nutrition. *Compr. Rev. Food Sci. Food Saf.* **9**, 417–436 (2010).
46. Rome Declaration on World Food Security and World Food Summit Plan of Action (FAO, 1996).
47. FAOSTAT: *Suite of Food Security Indicators* (FAO, 2022); <https://www.fao.org/faostat/en/#data/FS>
48. Department of Economic and Social Affairs *World Population Prospects* (United Nations, 2019); <https://population.un.org/wpp>
49. *The State of Food Insecurity in the World* (FAO, 2000).
50. *Global Food Security Index: Exploring Challenges and Developing Solutions* (Economist Impact, 2021); <https://impact.economist.com/sustainability/project/food-security-index/>

Acknowledgements

We thank the National Key Research and Development Program of China (2017YFD0200108) and the National Natural Science Foundation of China (31972520) for providing financial support to Y.M., M.F. and L.Q. We thank the Cornell Institute for Digital Agriculture (CIDA RIF2019) for supporting D.W. and AI-CLIMATE (NIFA-2023-67021-39829) for supporting J.L. on this project. We thank the China Scholarship Council for providing funds to Y.M. to pursue his study at Cornell University. We thank L. M. Johnson from the Cornell Statistical Consulting Unit for supporting the data analysis work. We thank all agricultural scientists and extension personnel who provided local technical assistance and management during the project.

Author contributions

Y.M., D.W., M.F. and J.L. designed the research; Y.M., D.W., M.F., L.Q., R.L. and J.L. collected field and other research data; Y.M., D.W., M.F. and J.L. contributed to methodology and data analysis; Y.M. and D.W. contributed to visualization; D.W., M.F. and J.L. supervised the work; Y.M. and D.W. wrote the first manuscript; Y.M., D.W., M.F. and J.L. edited the manuscript; all authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at
<https://doi.org/10.1038/s41561-023-01302-3>.

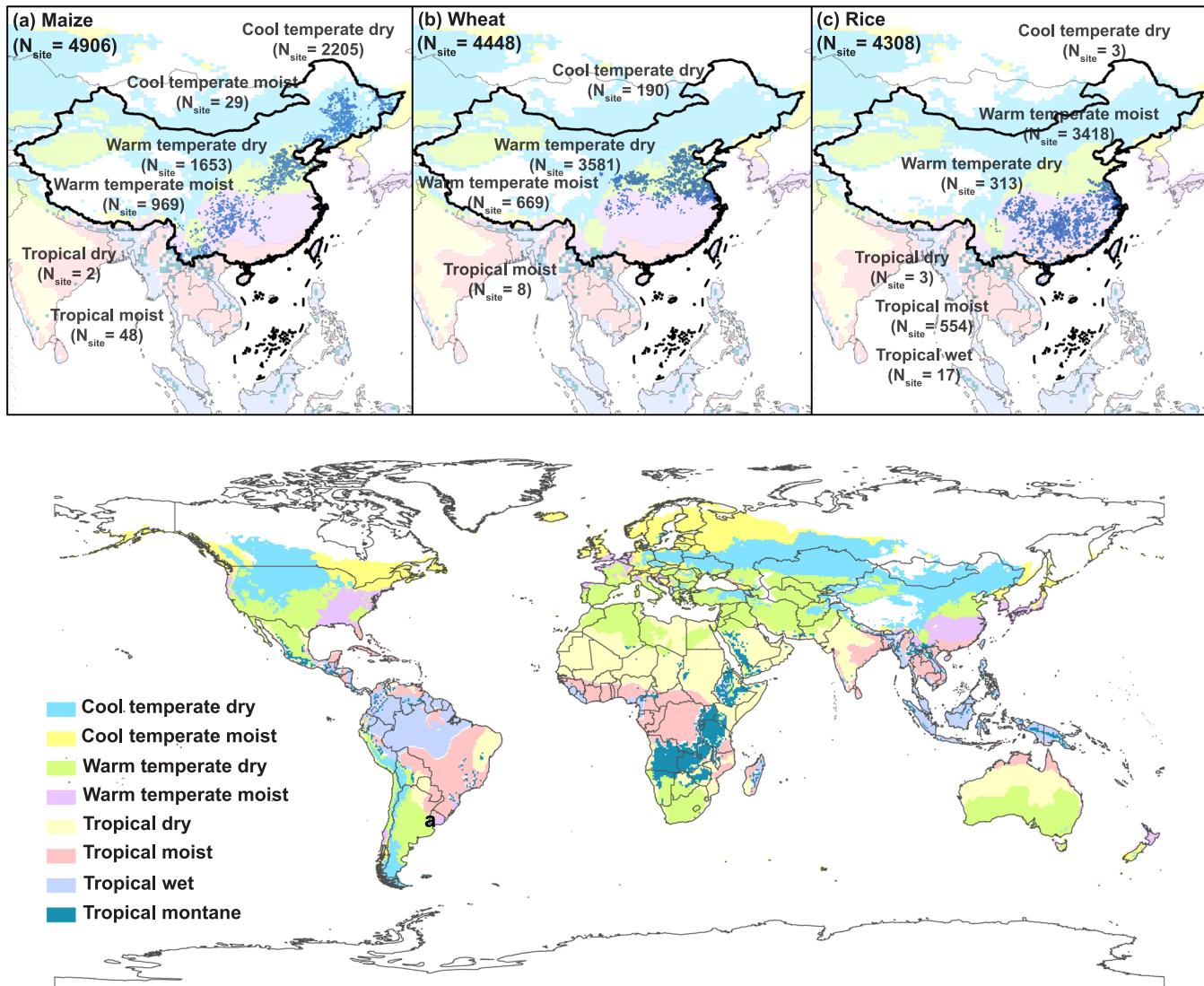
Supplementary information

The online version contains supplementary material available at
<https://doi.org/10.1038/s41561-023-01302-3>.

Correspondence and requests for materials should be addressed to Mingsheng Fan or Johannes Lehmann.

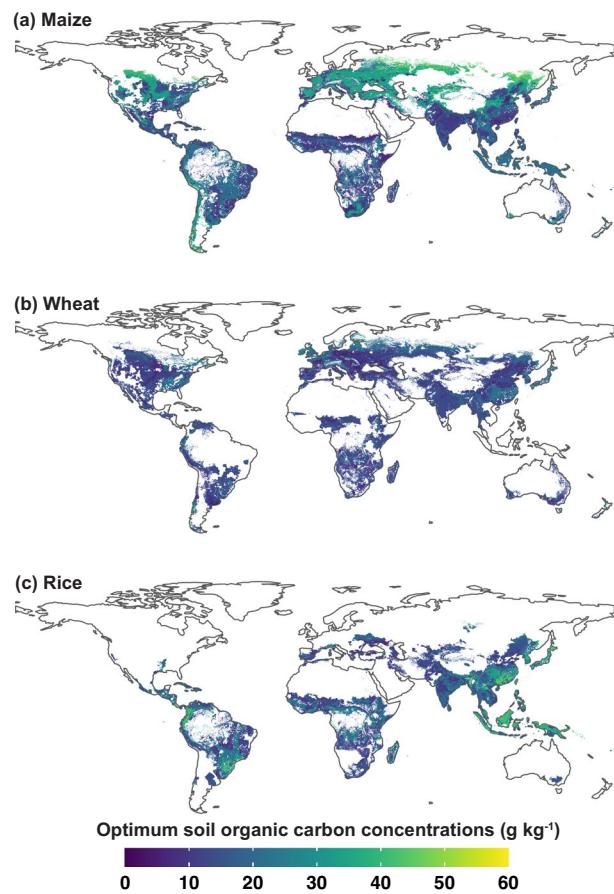
Peer review information *Nature Geoscience* thanks Marcel van der Heijden and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Xujia Jiang, in collaboration with the *Nature Geoscience* team.

Reprints and permissions information is available at
www.nature.com/reprints.



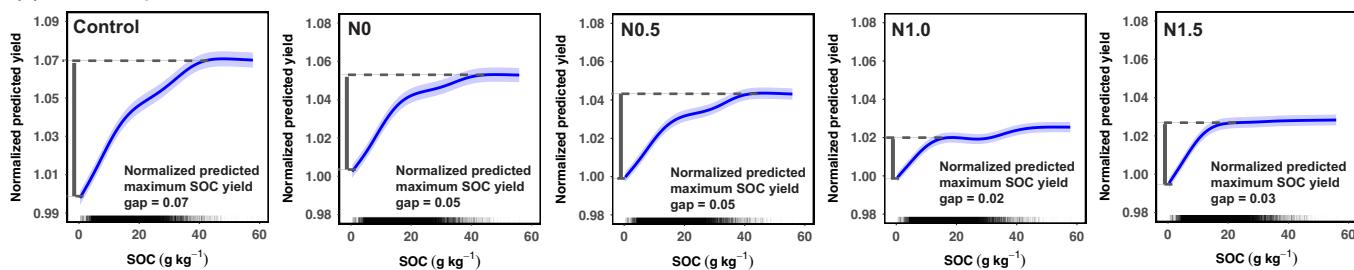
Extended Data Fig. 1 | Distribution of field trials for maize, wheat and rice.
a–c, Maize (a), wheat (b) and rice (c). The round symbols on the three top panels represent the locations of field trials. The bottom panel indicates the eight IPCC climate zones (large fractions of climate zones where the three studied crops distribute are covered by field trials). Numbers in brackets under the climate

name indicate the number of field trials for each climate zone of each crop. Numbers in brackets under the crop name indicate the total number of field trials for each crop. Basemap reproduced from Resource and Environment Science and Data Center, Institute of Geographic Sciences and Natural Resources, Chinese Academy of Sciences.

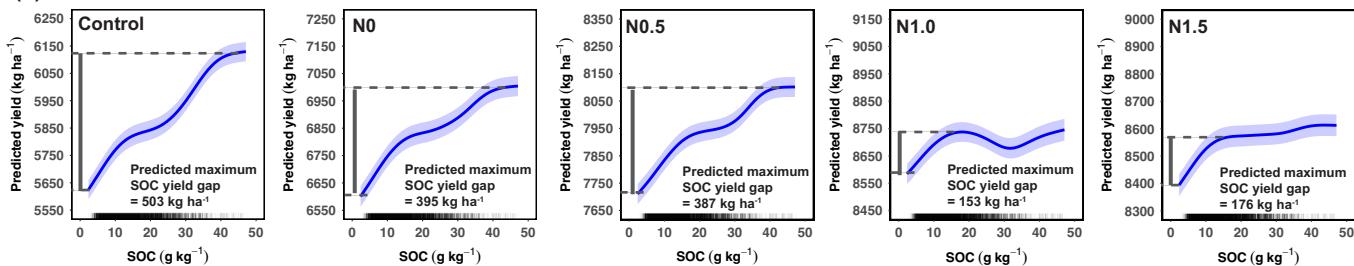


Extended Data Fig. 2 | Distribution of climate-and-soil specific optimum SOC for maize, wheat and rice. a–c, Maize (a), wheat (b) and rice (c). The optimum SOC is the value above which no additional increase in crop yield can be achieved by increasing SOC. Basemap reproduced from ArcGIS/Esri.

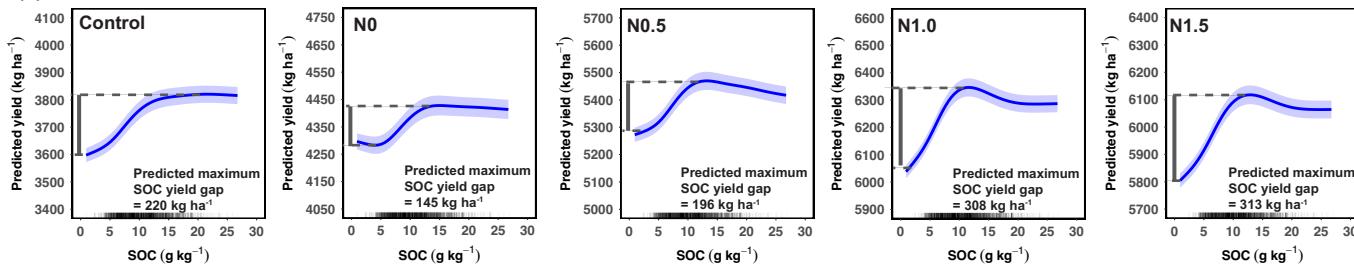
(a) Three crops combined



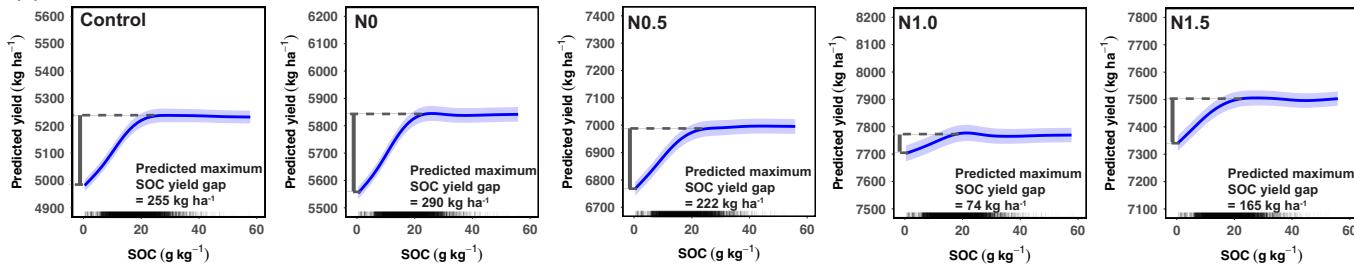
(b) Maize



(c) Wheat

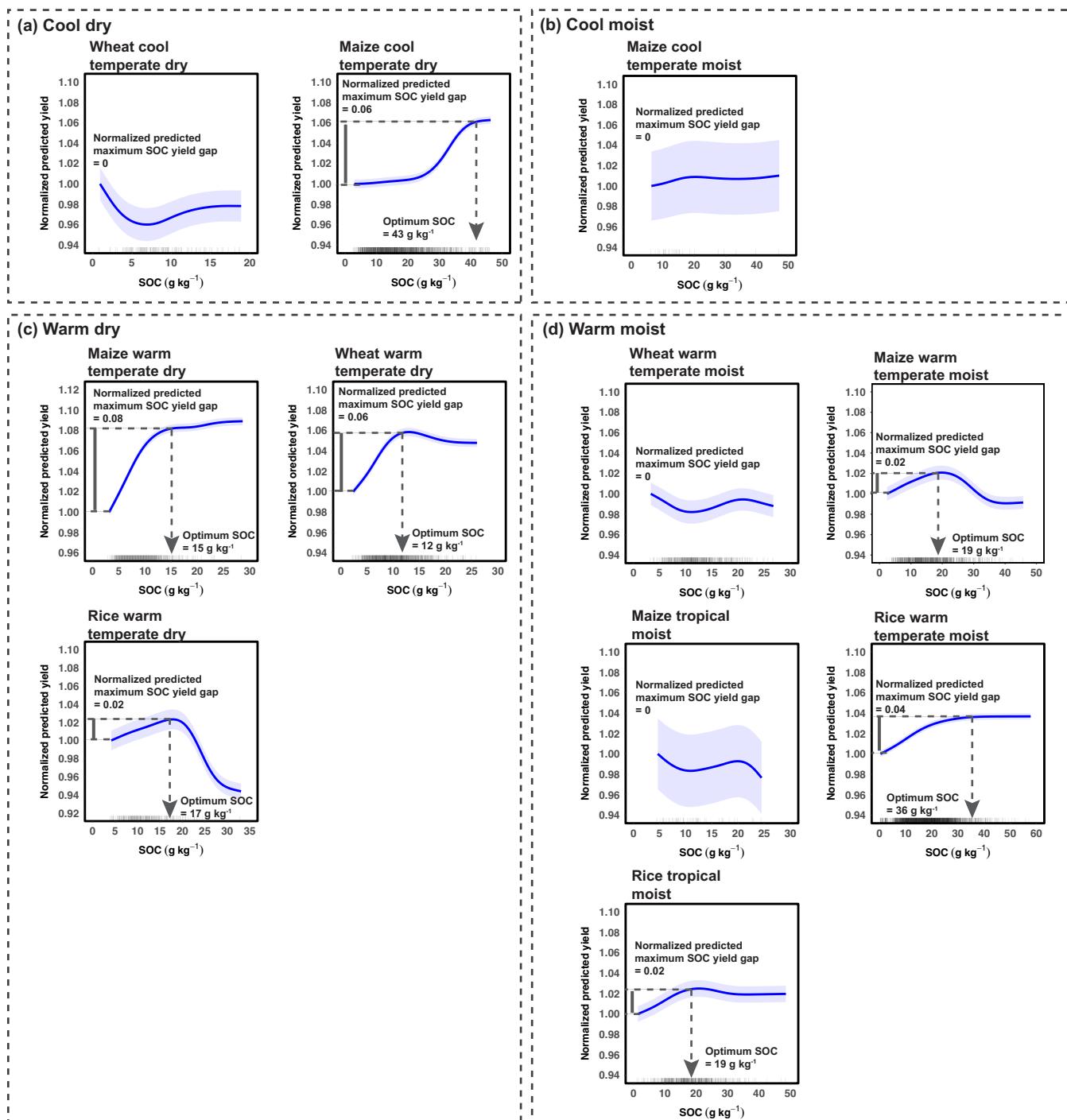


(d) Rice



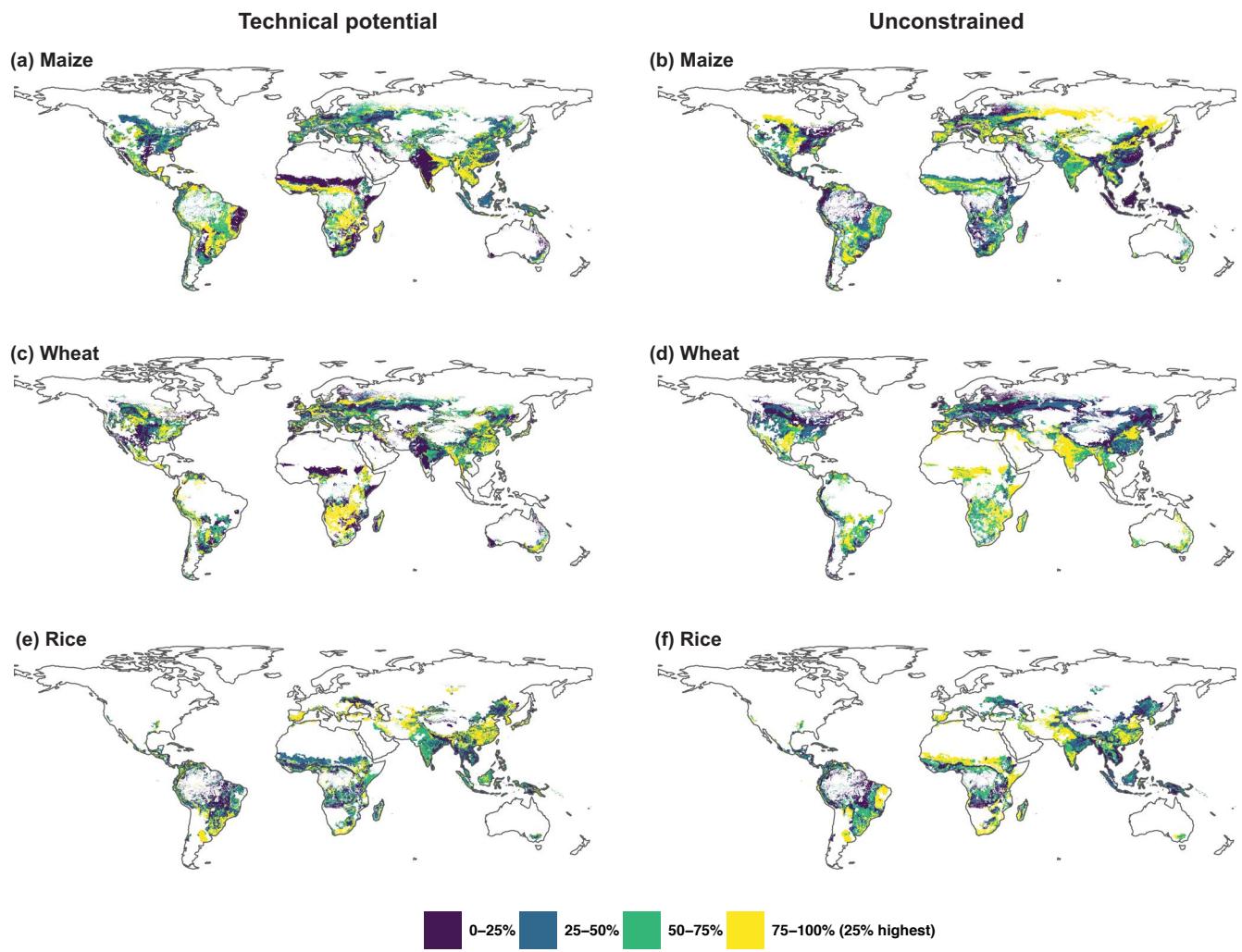
Extended Data Fig. 3 | Causal effect of soil organic carbon (SOC) on crop yield under different nitrogen input rates. **a**, three crops combined. **b**, maize. **c**, wheat. **d**, rice. The solid lines represent mean partial dependence of crop yield on SOC, and the ribbons represent 95% confidence intervals. N0, N0.5, N1.0 and N1.5 indicate 0, 50%, 100% and 150% of optimum nitrogen input rates, respectively, with optimum phosphorus and potassium input rates. Control

means no fertilizer inputs. For three crops combined, predicted yield was normalized as the ratio of on-farm measured yield in each trial to predicted minimum SOC yield gap for each crop under the same treatment. Predicted maximum SOC yield gap is the difference between predicted yield under optimum SOC and predicted minimum yield.



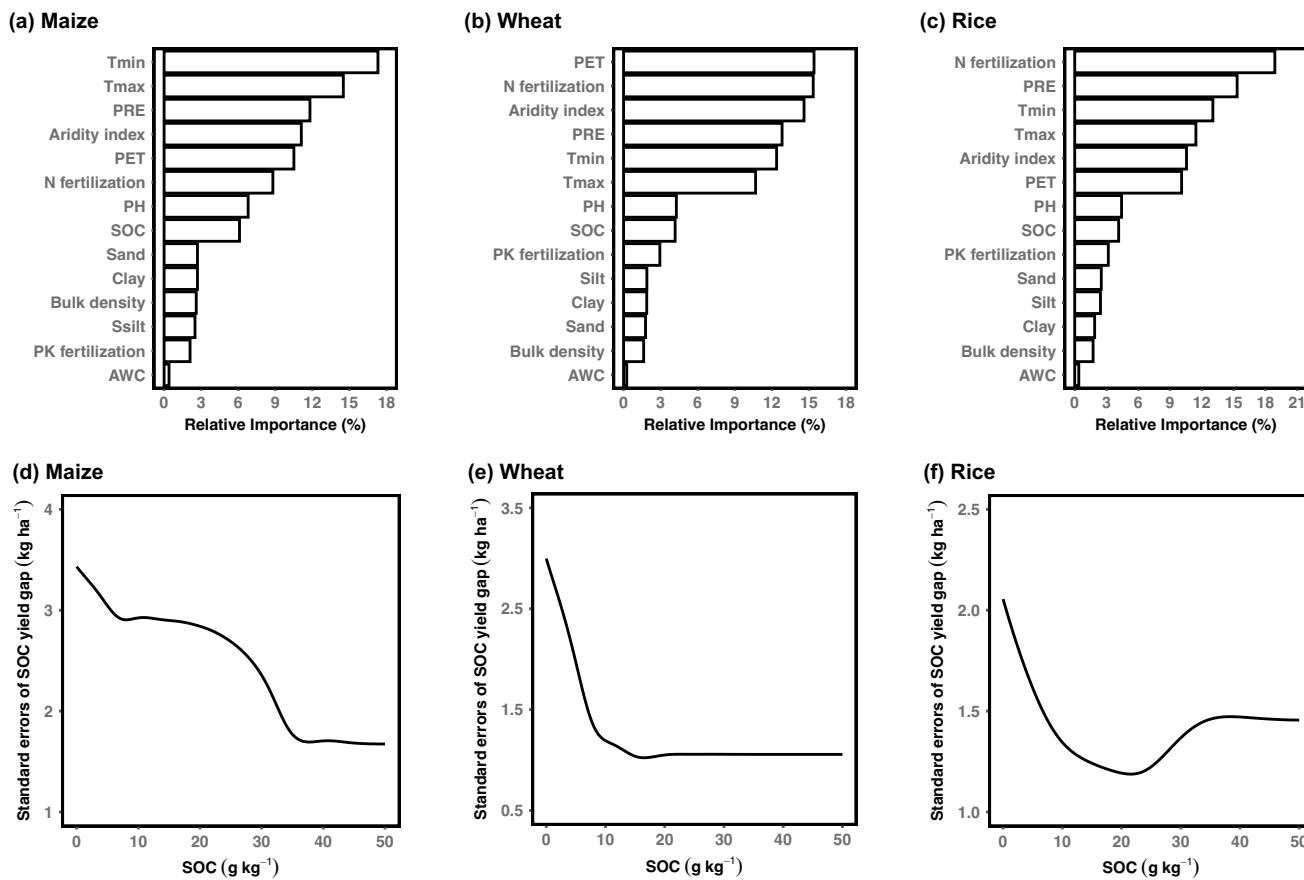
Extended Data Fig. 4 | Causal effect of soil organic carbon (SOC) on crop yield across climate zones for maize, wheat and rice. **a-d**, cool dry (a), cool moist (b), warm dry (c), and warm moist (d). The solid lines represent the mean partial dependence of normalized crop yield on SOC, and the ribbons represent 95% confidence intervals. Optimum SOC is the SOC level beyond which no significant ($p < 0.05$) yield increases were observed. The significance of the yield increase

was estimated by two-sided t-tests. Normalized predicted yield was calculated as the ratio of predicted yield from partial dependence to predicted minimum mean yield for each crop under each climate zone. Normalized predicted maximum SOC yield gap is the difference between normalized predicted yield under optimum SOC and normalized predicted minimum yield.



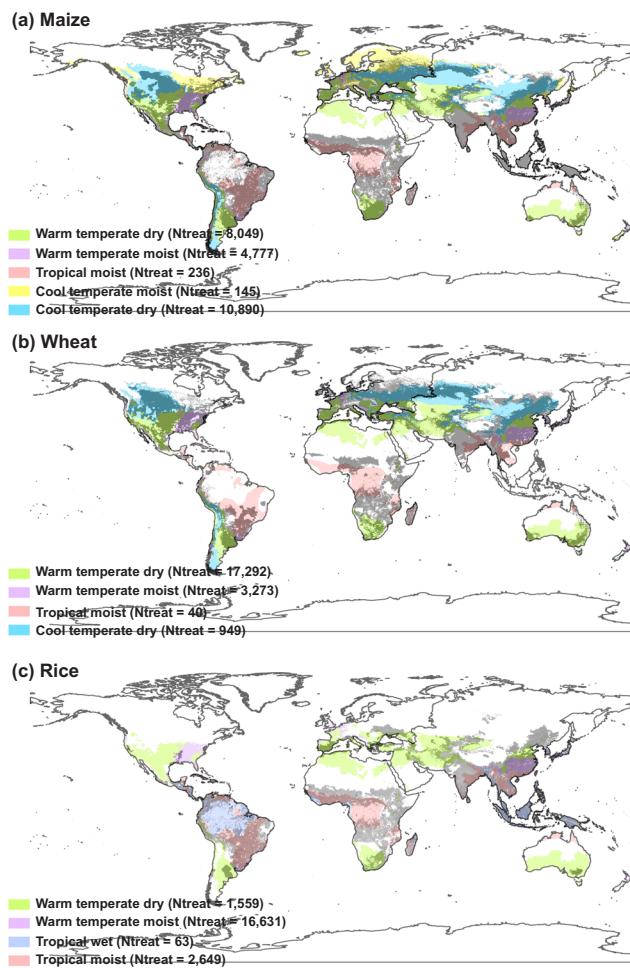
Extended Data Fig. 5 | Distribution of yield increases derived from soil organic carbon (SOC) for three studied crops. Regions on the maps are derived from Fig. 4 and classified into quartiles. **a, c, e**, technical potential represents yield gains when SOC is increased to current technical potential SOC levels for

maize (a), wheat (c), and rice (e). Current technical potential SOC levels can be achieved using a combination of high organic matter inputs and zero tillage; **b, d, f**, unconstrained represents yield gains when SOC increased to optimum levels for (b), wheat (d), and rice (f). Basemap reproduced from ArcGIS/Esri.



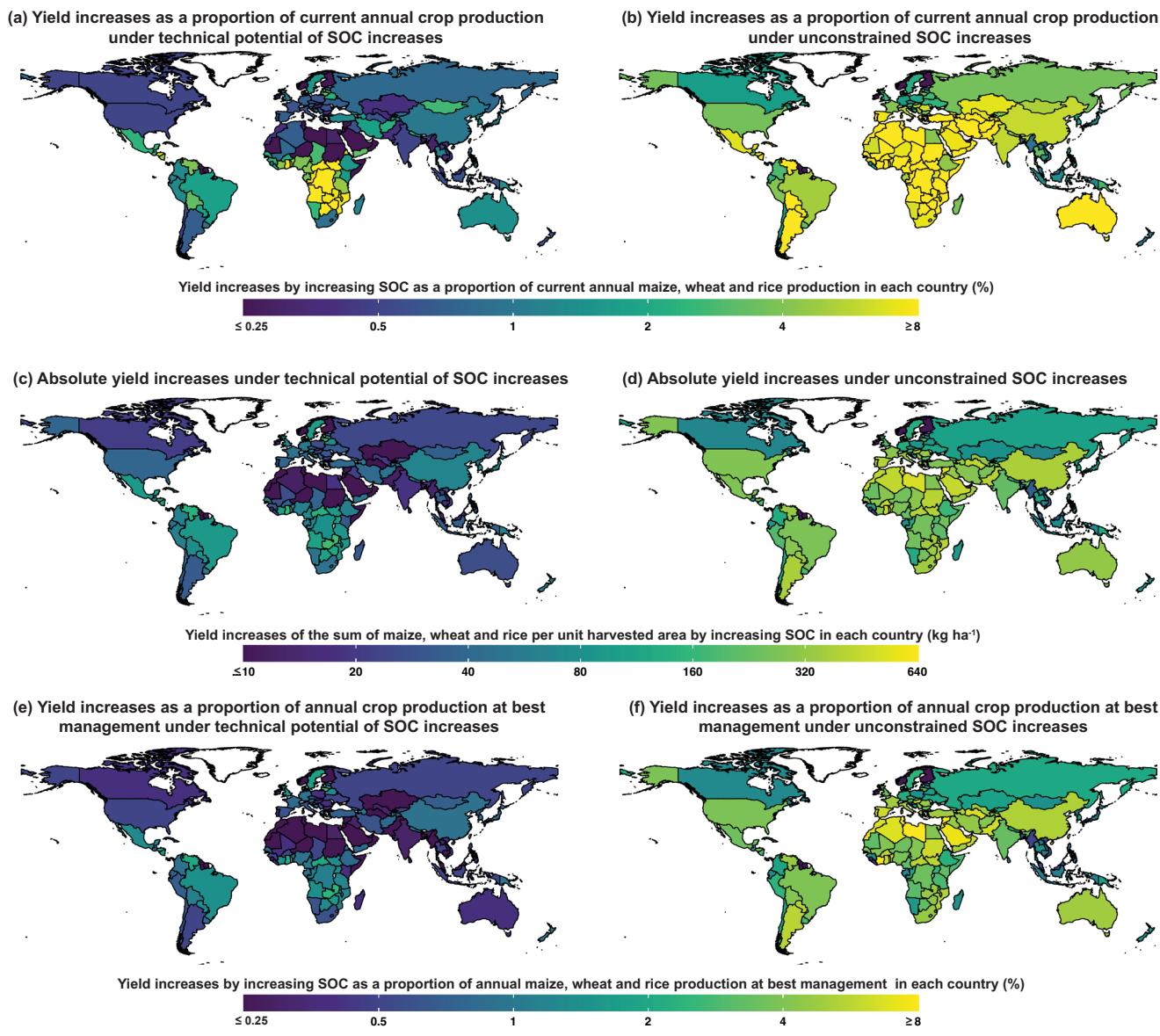
Extended Data Fig. 6 | Relative importance (%) of variables for the global extrapolation of yield models and its predicted standard errors for each model. a,b,c, bar charts indicate relative importance (permutation method; see Methods) of variables as predictors of random forest extrapolation yield models. d,e,f, line charts represent the predictions of standard errors for maize (d), wheat

(e), and rice (f) estimated by the jackknife method (See methods; Uncertainty analysis of global extrapolations). NPK fertilization = nitrogen, phosphorus and potassium fertilizer rate; PET = potential evapotranspiration; PRE, Tmax and Tmin = precipitation, maximum mean temperature and minimum mean temperature; AWC = available water capacity.



Extended Data Fig. 7 | Correspondence of distributions of climate zones where the majority of field trials are located and producing regions of the three studied crops. a, maize, b, wheat, c, rice. Climate zones were divided based on IPCC criteria⁴³. N_{treat} is the number of treatments rather than the number of field trials in each climate zones. The distributions of maize, wheat and rice

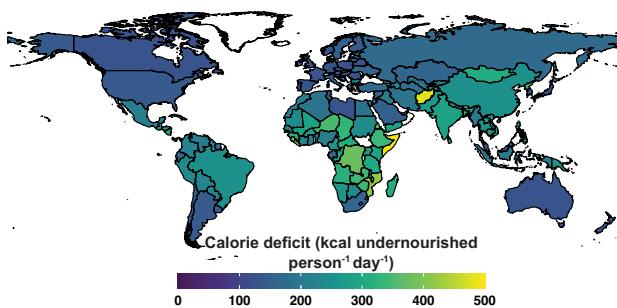
producing regions were obtained from EarthStat⁴⁴. The shaded regions are the global producing regions for the three studied crops. Climate zones where only few field trials distribute are excluded from the map including maize tropical dry zone ($N_{\text{treat}} = 10$), rice tropical dry zone ($N_{\text{treat}} = 15$) and rice cool temperate dry zone ($N_{\text{treat}} = 15$). Basemap reproduced from ArcGIS/Esri.



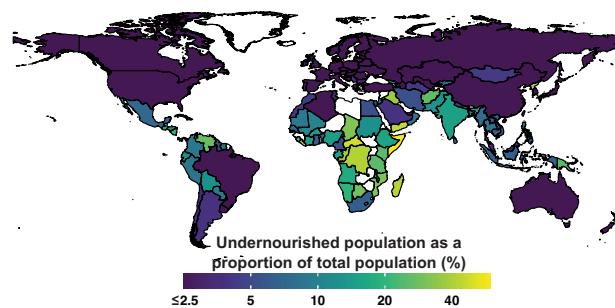
Extended Data Fig. 8 | Global yield increases of maize, wheat and rice by increasing soil organic carbon (SOC) in each country. Yield increases of the three crops by increasing SOC in each country was normalized to total yield increases of the three crops from SOC relative to annual current crop production of the three crops (a, b), to the yield increases of the three crops from SOC per unit harvest area (c, d), and to the total yield increases of the three crops from SOC relative to annual crop production of the three crops at best management (e, f). Total yield increase relative to annual current crop production was calculated by using yield increases of the three crops by increasing SOC in each country under technical potential and unconstrained increase in SOC divided

by its current annual crop production of the three crops in each country. Yield increases of the three crops per unit harvested area was calculated by using the yield increase of the three crops by increasing SOC in each country under the technical potential and unconstrained SOC increase divided by the total harvested area of the three crops in each country. Total yield increases relative to annual crop production at best management was calculated by using yield increases of the three crops by increasing SOC in each country under technical potential and unconstrained increase in SOC divided by annual crop production of the three crops at best management in each country.

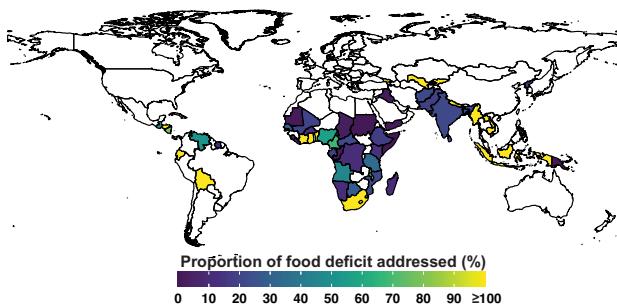
(a) Depth of food deficit



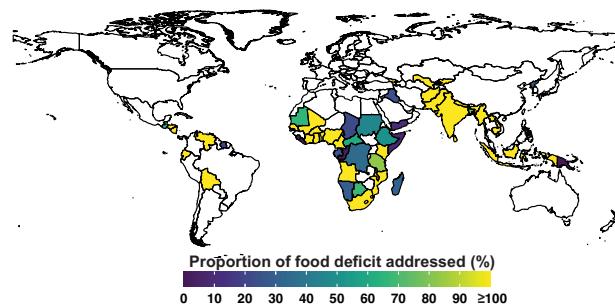
(b) Prevalence of undernourishment



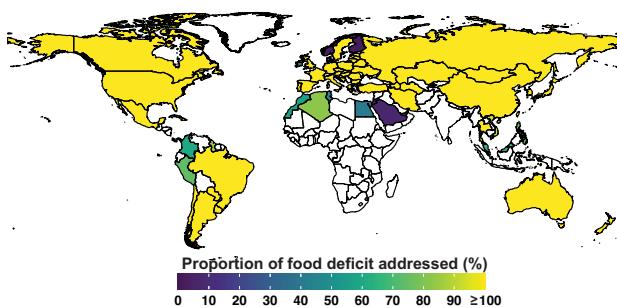
(c) Technical potential to address food deficit in countries with weak and moderate food security environment



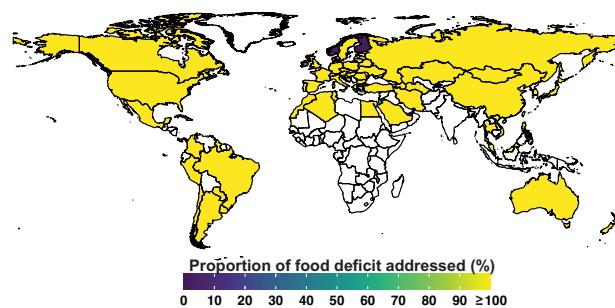
(d) Unconstrained to address food deficit in countries with weak and moderate food security environment



(e) Technical potential to address food deficit in countries with good and very good food security environment



(f) Unconstrained to address food deficit in countries with good and very good food security environment



Extended Data Fig. 9 | Depth of food deficit, prevalence of undernourishment and the proportions of food deficit addressed by improving SOC under the technical potential or unconstrained SOC increases. Depth of food deficit (a) is the average calorie deficiency per capita per day below the minimum dietary daily energy requirement in each country. Prevalence of undernourishment (b) is the ratio of total undernourished population to the total population in each country that experiences national food deficits. Proportions of food deficit addressed were estimated by using the kilocalories provided by a daily increase of maize, wheat and rice per undernourished person resulting from a

SOC increase under either the technical potential (c,e) or unconstrained SOC increases (d,f) for each country divided by the daily food deficit in kilocalories per undernourished person (Supplementary Table 6). The countries (c-f) were grouped based on the Global Food Security Index⁵⁰ and classified as ‘weak and moderate food security environment’ or ‘good and very good food security environment’ (Supplementary Table 6). The technical potential represented yield gains when SOC increased to levels that are achieved using a combination of high organic matter inputs and zero tillage; an unconstrained increase in SOC represented yield gains when SOC increased to optimum levels.