

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used for data collection.
Data analysis	We resampled global raw data using Arcgis and did all analysis in Stata version 16.0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

We utilized nitrogen budget data at a 0.5°x0.5° spatial resolution, including fertilizer input, biological nitrogen fixation, deposition, and manure, from the IMAGE-GNM database (<https://doi.org/10.5194/gmd-8-4045-2015>). Cropland data, measured in square kilometers per grid at the same resolution, was obtained from the History Database of the Global Environment (HYDE v3.2) (<https://doi.org/10.5194/essd-9-927-2017>). Rural population data was sourced from the WorldPop dataset (<https://www.worldpop.org/>). Soil physical properties, including bulk density, pH, and organic carbon density, were extracted from the surface layer (0–30 cm) at a

1 km×1 km resolution (ISRIC dataset <https://files.isric.org/>). Phosphorus content was sourced from (<https://doi.org/10.6084/m9.figshare.14583375>). Biotic data, covering fungal richness, bacterial richness, viral abundance, nematodes, and worm abundance, were obtained from: Tedersoo et al. (2014) (<https://www.science.org/doi/10.1126/science.1256688>), Delgado-Baquerizo et al. (2018) (<https://www.science.org/doi/10.1126/science.aap9516>), Ma et al. (2024) (<https://www.nature.com/articles/s41559-024-02347-2>), van den Hoogen et al. (2019) (<https://www.nature.com/articles/s41586-019-1418-6>), Phillips et al. (2019) (<https://www.science.org/doi/10.1126/science.aax4851>), respectively. Soil pesticide data, including fungicides, herbicides, and insecticides, was sourced from Tang et al. (2021) (<https://www.nature.com/articles/s41561-021-00712-5>). Climate data for 2020, including temperature and precipitation at a 0.5°×0.5° resolution, was obtained from CRU TS v4.06 (<https://doi.org/10.1038/s41597-020-0453-3>). For scenario analysis, we collected cropland area, population, temperature, and precipitation data under three SSP-RCP scenarios (SSP1-RCP2.6, SSP2-RCP4.5, and SSP5-RCP8.5). Cropland area projections until 2050 were sourced from the LUH2 dataset (<https://luh.umd.edu/data.shtml>). Population projections until 2050 were obtained from gridded datasets for population and economy under SSPs (<https://doi.org/10.12006/j.issn.1673-1719.2022.106>). Monthly temperature and precipitation data (2015–2050) at 0.9°×1.25° resolution were sourced from the World Climate Research Programme Coupled Model Intercomparison Project (CMIP6) (<https://esgf-node.llnl.gov/projects/cmip6/>). Data supporting this study's findings are included in the article and Supplementary Information. Additional data are sourced from publicly available databases and literature, as cited in the reference list. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender N/A

Reporting on race, ethnicity, or other socially relevant groupings N/A

Population characteristics N/A

Recruitment N/A

Ethics oversight N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	In this study, we establish a comprehensive global soil health dataset at a 0.5 × 0.5-degree resolution for the world's croplands. This dataset encompasses not only conventional factors such as soil pH, nutrients, and bulk density but also accounts for soil bacteria, viruses, and fauna (See more details in Table S1). To quantify the contribution of soil health to agricultural productivity, our dataset also integrates aspects of agricultural management and climatic conditions—factors of significant interest in prior research through comprehensive statistical analysis, including structural equation model and multiple regression model. Beyond assessing crop yields, we employ NUE as a metric to signify the environmental impact of agricultural practices: the higher the NUE, the lower the detrimental environmental implications. Finally, we analyze the stewardship of soil health and other interrelated components in the context of global crop production based on the scenario analysis.
Research sample	This study focuses on soil and nitrogen dataset at a 0.5 × 0.5-degree resolution for the world's croplands.
Sampling strategy	Sampling strategy is no relevant to our study.
Data collection	All data are from government websites, published papers and publicly database.
Timing and spatial scale	Timing scale: Around year 2020 Spatial scale: 0.5 × 0.5-degree resolution
Data exclusions	We corrected the raster data from the IMAGE model based on the nitrogen budget of the CHANS model. Some raster values exceeded reasonable limits, and we set them to the upper limit of the reasonable range. The upper limit was determined based on Supplementary Table 16.
Reproducibility	Our data are available on model is simply designed, everyone can reproduce the result.

Randomization

Randomization is no relevant to our study.

Blinding

Blinding is not relevant to this study, as no treatment is applied.

Did the study involve field work?

☐

Yes

☒

No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A