

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	Downloaded data was processed using R 4.3.2 with the following packages: tidyverse 2.0.0, sf 1.0-15, terra 1.7-71, exactextractr 0.10.0, and LingrNR 0.1.0 for yield modelling. All codes for data generation are in the replication repository.
Data analysis	We used R 4.3.2 for all analyses. We used the grf 2.3.2 package for causal forest estimation. All custom codes for data analysis and figures are available in the replication repository. Maps and figures were created with ggplot2 3.5.0.

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](#)

R-codes, preprocessed data, and model outputs to replicate figures and tables are permanently available at <https://phenoroam.phenorob.de/> with the identifier "254e9ed1-bb57-42c6-83bf-851de7ed11f7". For a detailed overview of data sources see Table S1. The plant species estimations are currently not publicly available;

for inquiries contact the authors. A sample of the species richness dataset for 2021 is available at <https://doi.pangaea.de/10.1594/PANGAEA.967052>. Field shapes are available at <https://zenodo.org/records/10619783>. The indicator of mowing frequency is available at <https://doi.org/10.5281/zenodo.5571613>; the database of protected areas at www.protectedplanet.net; topography at <https://doi.org/10.5523/bris.s5hqmjcdj8yo2ibzi9b4ew3sn>; land cover at <http://doi.org/10.1594/pangaea.910837>; weather and climate data at https://opendata.dwd.de/climate_environment/CDC/; soil grids at <https://soilgrids.org>; soil hydraulic properties at <https://esdac.jrc.ec.europa.eu/resource-type/datasets>; soil depth at http://daac.ornl.gov/cgi-bin/dsvviewer.pl?ds_id=1304; and agricultural census data at <https://www.regionalstatistik.de/genesis/online>.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	Not applicable.
Population characteristics	Not applicable.
Recruitment	Not applicable.
Ethics oversight	Not applicable.

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](https://www.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	We used an observational panel dataset sourced via remote sensing and secondary data sources to estimate counterfactual plant species richness using generalized random forests.
Research sample	Our data contains the full population of agricultural grassland fields in Germany, as identified via secondary data. Our predicted outcome variable, plant species richness, is spatially restricted to its area of applicability. Our results are representative for grasslands larger than 0.3 hectare mown between 1 and 5 times per year between 2017 and 2020 (including normal and dry years).
Sampling strategy	For computational efficiency and to avoid spatial autocorrelation, we used a simple random sample of 50,000 observations.
Data collection	We extracted secondary data using the sf package in R in combination with terra and exactextractr for raster data.
Timing and spatial scale	For each field, we obtained annual estimates of mowing frequency, plant species richness and covariates between 2017 and 2020. Note, that our outcome is plant species richness in the subsequent year, so these values refer to 2021.
Data exclusions	We excluded fields smaller than 0.3 hectare and those not overlapping with the mowing event and species richness rasters.
Reproducibility	All codes are in the replication repository. We used random seeds to allow sample replication.
Randomization	We controlled for non-random treatment status using topographic, climatic, soil, economic, and landscape indicators measured at different spatial and temporal resolutions. We used causal forest, a doubly-robust approach that to some extent captures unobserved variation associated with treatment status.
Blinding	Blinding was not necessary, because no randomization was done.

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

Methods

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