

Reporting Summary

Nature Research 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 Research policies, see [Authors & Referees](#) 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 (n) 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
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- 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. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- 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 d , Pearson's r), 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

All datasets used in this the current study were fetched from other sources and downloaded online from their respective sources.

Data analysis

Data analysis was executed in Python 3.6, with the custom code available at: <https://github.com/charlottegiseleweil/ResilienceOfNutrition>. Modeling was primarily supported by the Python library xgboost 0.9.0. More details about the environment of analysis are available in the repository's readme file.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [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 list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated during and/or analysed during the current study are available in the Zenodo repository: <https://zenodo.org/record/3574974#.XipntxNKjOR>.

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.

Ecological, evolutionary & environmental sciences study design

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

Study description

Part (1) Modeling caloric yields

Aggregated caloric yield (of 100 nutritionally relevant crops, in calories per hectare) were modeled based on input variables related to climate, soil and management practices, as described in the first two section of Methods.

Part (2) Caloric Sufficiency analysis

Caloric sufficiency was calculated based on projected production and demand, joining our modeled results to existing projections of the SSPs (Shared Socio-Economic Pathways). Details about the methodology can be found in the third section of Methods (Caloric sufficiency calculation).

Research sample

The current study is based on existing datasets, listed in Methods.

Sampling strategy

The sampling for holding out a validation subset of the data, and to ensure the validity of aggregating all crops' caloric yields, were done randomly.

Data collection

All datasets used in this the current study were fetched from other sources and downloaded online from their respective sources.

Timing and spatial scale

Part (1) Modeling caloric yields

All datasets used to build and train the model were chosen circa 2000, as close as possible to this temporal scale with data availability constraints, to be consistent with the response variable (the yields dataset from Monfreda, 2008). Similarly for spatial resolution, all datasets were directly taken or re-sampled at 5arcmin, as the resolution of the the yields dataset. Spatial and temporal scales of these datasets are detailed in the first section of Methods.

Part (2) Caloric Sufficiency analysis

To explore caloric sufficiency in more current food security context, we link our results with most recent available relevant information (2000–2018), at national scales. Spatial and temporal scales of these datasets are detailed in the last section of Methods.

Data exclusions

No data were excluded from the analysis.

Reproducibility

At each main step of the analysis, calculation were replicated independently to cross-validate.

Randomization

When building the yields model (part 1), we hold out a random set of 10% pixels as validation, the random seed was fixed to 123. (See code for details)

To ensure the validity of aggregating all crops' caloric yields, we sampled 4 million pixels pairs (with a set of 2000 randomly selected pixels, each associated with 2000 others). This was done repetitively with different samples of pixel pairs, to verify that the validity of sampling strategy. (See code and Figure S1). The results reported (person correlation) were found unchanged up to the second decimals, when trying several other pixel pairs samples.

Finally, to construct the clusters leading to figure S2, we fixed the random seed at 0.

Blinding

Blinding is not applicable to this study.

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
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging