Short description of each input dataset

The datasets used in the analysis are compiled in Table 1. In the following, the data will be presented briefly. For a detailed description of the methodology, refer to the respective papers.

Crop yield data for maize, rice, soybean, and wheat were taken from "SPAM2010" (Spatial Production Allocation Model) datasets presented by Yu et al. (2020). The maps are produced by disaggregating non-spatial crop statistics across farming systems and then allocate these values to a spatial grid using an optimization process based on multiple yield-influencing statistics. In the analysis the yield in kg per hectare by grid cell and the area in hectare per grid cell across all production systems are used.

Table 1: GLM Input datasets

Dataset	Definition	Spatial resolution	Temporal reference	Source	Available online
SPAM	Yield (kg/ha), harvested area (ha/cell)	5 arcmin	2010	Yu et al. (2020)	https://doi.org/10.7910/DVN/ PRFF8V
GAEZ v4 AEZ Factors	Thermal regime class, Moisture regime class, Soil/terrain related classes	5 arcmin 5 arcmin 30 arcsec	2010	Fischer et al. (2021)	https://gaez.fao.org/pages/data-viewer
PEST-CHEMGRIDS	Application rate (kg/ha) of 20 active ingredients for 10 dominant crops and 4 aggregated crop classes	5 arcmin	2015	F. Maggi, Tang, La Cecilia, and McBratney (2020)	https://doi.org/10.7927/weq9-pv30
Global Map of Irrigation Areas - Version 5	Area equipped for irrigation (% of total area)	5 arcmin	2005	Siebert, Henrich, Frenken, and Burke (2013)	https://data.apps.fao.org/map/catalog/srv/api/records/f79213a0-88fd-11da-a88f-000d939bc5d8
Gridded nitrogen and phosphorus fertilizer use	N and P application rate (g/m²)	0.5◦degree	1900-2013	Lu and Tian (2016)	https://doi.pangaea.de/10.1594/PANGAEA.863323
Global gridded dataset of manure nitrogen production and application	N manure application (kg/km²)	5 arcmin	1860-2014	Zhang et al. (2017b)	https://doi.pangaea.de/10.1594/PANGAEA.871980
A global gridded data set on tillage (V. 1.1)	Six tillage systems (dominant system/cell)	5 arcmin	Around 2005	Porwollik, Rolinski, and Müller (2019)	https://doi.org/10.5880/PIK.2019.009

As discussed above, only high-level *climatic variables* are included as the goal of the prediction is to assess the lack of anthropogenic inputs without climatic changes. The Agroecological Zones (AEZ) classification is a part of the global AEZ methodology (version 4) and is introduced as a framework for broad-scale analysis and planning (Fischer, 2021). The AEZ classes are defined by a combination of temperature regime, moisture regime and soil/terrain related classes. The soil/terrain related classes include an irrigation class which renders redundant information as the actually irrigated area per cell was incorporated into the analysis as a separate variable (see below). Therefore, instead of the combined AEZ classes, the temperature regime, the moisture regime, and the soil/terrain related classes are included as separate variables. This ensures that only the soil/terrain related classes contain some redundant information within the irrigation category and limits the number of total categories. While temperature and moisture class datasets are available at five arcmin resolution, the soil/terrain dataset is provided at 30 arcsec resolution. To fit the target resolution, it was downsampled to five arcmin. Table 2 gives an overview over the individual classes of the temperature regime, the moisture regime, and the soil/terrain related conditions.

Table 2: List of AEZ temperature regime, moisture regime and soil/terrain related classes (Fischer, 2021). LGP = Length of Growing Period

Temperature Regime Classes	Moisture regime classes	Soil/terrain related classes
TRC1=Tropics, lowland	M1=LGP < 60days	S1=Dominantly very steep terrain
TRC2=Tropics, highland	M2=LGP 60-120 days	S2=Dominantly hydromorphic soils
TRC3=Subtropics, warm	M3=LGP 120-180 days	S3=No or few soil/terrain limitations
TRC4=Subtropics, moderately cool	M4= LGP 180-225 days	S4= Moderate soil/terrain limitations
TRC5=Subtropics, cool	M5= LGP 225-270 days	S5=Severe soil/terrain limitations
TRC6=Temperate, moderate	M6= LGP 270-365 days	L1=Water
TRC7=Temperate, cool	M7= LGP 365+ days	L2=Built-up/Artificial
TRC8=Boreal/Cold, no permafrost		L3=Irrigated soils
TRC9=Boreal/Cold, with permafrost		
TRC10=Arctic/Very cold		

For *irrigation*, the fraction of actually irrigated cropland per cell was calculated based on the “Global Map of Irrigation Areas - Version 5” (Siebert et al., 2013) by multiplying the area equipped for irrigation in hectares per cell with the area actually irrigated as percentage of the area equipped for irrigation per cell. The irrigated areas were determined using subnational statistics derived from national census data or irrigation sector studies. The subnational unit values were then allocated to grid cells by aligning regional and local irrigation maps from various sources with satellite imagery on irrigation areas or other related factors if irrigation data were missing (Siebert et al., 2013). They report that layers for areas equipped for irrigation are much more reliable than layers on actually irrigated areas and water sources because the alignment process was only applied to the layers containing areas equipped for irrigation.

The “PEST-CHEMGRID” dataset provides spatially explicit layers of the application rates of 20 most used active ingredients in kg per hectare per cell for six dominant crops and four aggregated crop classes (F. Maggi et al., 2020). It is based on the USGS Pesticide National Synthesis Project (USGS/PNSP) which compiled the application mass of all active ingredients for 48 US States. Emanating from this dataset, application rates for the 20 most important pesticides were calculated for the US in the base year of 2015. Together with 20 globally gridded environmental quantities the US rates were used in a polynomial extrapolation to estimate values for the rest of the world. Afterwards the data were corrected for local pesticide restrictions (EU) and the usage of genetically modified (GM) crops and harmonized against the FAOSTAT country-level values. For the analysis a *pesticide variable* for each crop was compiled by summing up the active ingredients in each cell. The upper bounds of the range for each active ingredient were chosen to remain conservative for the catastrophe yield predictions. Federico Maggi, Tang, La Cecilia, and McBratney (2019) state that the selected ingredients account for 84.2% of pesticide usage in the US in 2015.

Nitrogen (N) fertilizer application rates were taken from the most recent available year of the "Half-degree gridded nitrogen and phosphorus fertilizer use for global agriculture production during 1900-2013" dataset compiled by Lu and Tian (2016). They used total country application mass provided by the International Fertilizer Association (IFA) and the Food and Agriculture Organisation (FAO) and harvested area per grid cell derived from the M3-crop data to calculate tabulated application rates. The tabulated data were then interpolated to achieve a half-degree gridded distribution and subsequently harmonized with IFA totals (Lu & Tian, 2017). As the half degree resolution does not align with the other datasets, the layers were upsampled to five arcmin resolution. The unit of the application rates are given as g per m² which was converted to kg per hectare to match the unit of the crop yield.

The "Global gridded dataset of manure nitrogen production and application" (Zhang et al., 2017b) supplies spatially explicit *manure production and application rates* in kg per km² per cell at five arcmin resolution. The manure production rate was calculated using the distribution of livestock provided by the Global Livestock Impact Mapping System (GLIMS), the animal-specific excretion rates from the Intergovernmental Panel on Climate Change (IPCC) 2006 guidelines and FAOSTAT data to trace yearly changes in livestock populations. The manure application rate was estimated based on the spatial distribution of different livestock management systems in different agroecological zones (Zhang et al., 2017a). As for the fertilizer application rate, the values were converted to kg per hectare to match the crop yield unit. The most recent available year was used for the analysis.

The N manure and N fertilizer application rate datasets presented above were summed up into a combined N total layer. This was done because the interest of the analysis lies with the effect reduced N input has on yield and not with the effect of N input from different sources. Moreover, it was taken as a measure to reduce the number of variables and possible multicollinearity between them.

Mechanization is the only selected factor which required the use of a proxy as no spatially explicit data on the degree of mechanisation in agriculture was available. The "global gridded data set on tillage (V. 1.1.)" created by Porwollik, Rolinski, and Müller (2019) served as a surrogate to determine if an area is farmed with motorized agricultural machinery or based on human and animal draft power. The authors classified tillage practices into six tillage systems according to literature findings, namely:

- 1 = conventional annual tillage
- 2 = traditional annual tillage
- 3 = reduced tillage
- 4 = Conservation Agriculture
- 5 = rotational tillage, and
- 6 = traditional rotational tillage.

The tillage systems were allocated to gridded cropland areas at five arcmin resolution based on spatial datasets for crop type, water management regime, field size, water erosion, income, and aridity (Porwollik, Rolinski, Heinke, & Müller, 2019). A large factor for the classification of tillage systems is the involvement of heavy machinery as it facilitates mixing soil in greater depth. Hence, it is possible to determine which systems rely on machinery for tilling and which do not by means of the decision tree and the characteristics of the tillage systems presented by Porwollik, Rolinski, Heinke, and Müller (2019). It is reasonable to assume other farm activities such as sowing and harvesting are also carried out with machinery if tilling is mechanized. Therefore, the tillage systems are reclassified into either 0 = non-mechanized or 1 = mechanized: tillage systems 2, 3 and 6 are classified as 0 and systems 1, 4 and 5 as 1. Conservation agriculture is classified as mechanized even though tillage is reduced to almost zero because currently conservation agriculture is most widely adopted in North and South America and Australia (Kassam, Friedrich, & Derpsch, 2019) where agriculture tends to be mostly mechanized. The dataset provides separate layers for 42 different crop types but in the analysis a combined layer is used so that a larger area is covered. It is assumed that regions with highly mechanized cropping systems for one crop generally do not farm other crops by hand or with animals.

References

Fischer, G. (2021). Global Agro-Ecological Zones v4 – Model documentation. FAO. https://doi.org/10.4060/cb4744en

Kassam, A., Friedrich, T., & Derpsch, R. (2019). Global spread of Conservation Agriculture. International Journal of Environmental Studies, 76(1), 29–51. https://doi.org/10.1080/00207233.2018.1494927

Lu, C., & Tian, H. (2016). Half-degree gridded nitrogen and phosphorus fertilizer use for global agriculture production during 1900-2013 [Data set]. In Supplement to: Lu, C; Tian, H (2017): Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: Shifted hot spots and nutrient imbalance. Earth System Science Data, 9(1), 181-192, https://doi.org/10.5194/essd-9-181-2017. PANGAEA. https://doi.org/10.1594/PANGAEA.863323

Maggi, F., Tang, F. H. M., la Cecilia, D., & McBratney, A. (2019). PEST-CHEMGRIDS, global gridded maps of the top 20 crop-specific pesticide application rates from 2015 to 2025. Scientific Data, 6(1), Article 1. https://doi.org/10.1038/s41597-019-0169-4

Porwollik, V., Rolinski, S., Heinke, J., & Müller, C. (2019). Generating a rule-based global gridded tillage dataset. Earth System Science Data, 11(2), 823–843. https://doi.org/10.5194/essd-11-823-2019

Siebert, S., Henrich, V., Frenken, K., & Burke, J. (2013). Update of the digital global map of irrigation areas to version 5. https://doi.org/10.13140/2.1.2660.6728

Yu, Q., You, L., Wood-Sichra, U., Ru, Y., Joglekar, A. K. B., Fritz, S., Xiong, W., Lu, M., Wu, W., & Yang, P. (2020). A cultivated planet in 2010 – Part 2: The global gridded agricultural-production maps. Earth System Science Data, 12(4), 3545–3572. https://doi.org/10.5194/essd-12-3545-2020

Zhang, B., Tian, H., Lu, C., Dangal, S. R. S., Yang, J., & Pan, S. (2017). Global manure nitrogen production and application in cropland during 1860–2014: A 5 arcmin gridded global dataset for Earth system modeling. Earth System Science Data, 9(2), 667–678. https://doi.org/10.5194/essd-9-667-2017