scientific **data**

Check for updates

DATA DESCRIPTOR

A global dataset for the projected impacts of climate change on four major crops

Toshihiro Hasegawa[1]✉, Hitomi Wakatsuki[1], Hui Ju[2], Shalika Vyas[3], Gerald C. Nelson[4], Aidan Farrell[5], Delphine Deryng[6], Francisco Meza[7] & David Makowski[8]

Reliable estimates of the impacts of climate change on crop production are critical for assessing the sustainability of food systems. Global, regional, and site-specific crop simulation studies have been conducted for nearly four decades, representing valuable sources of information for climate change impact assessments. However, the wealth of data produced by these studies has not been made publicly available. Here, we develop a global dataset by consolidating previously published meta-analyses and data collected through a new literature search covering recent crop simulations. The new global dataset builds on 8703 simulations from 202 studies published between 1984 and 2020. It contains projected yields of four major crops (maize, rice, soybean, and wheat) in 91 countries under major emission scenarios for the 21st century, with and without adaptation measures, along with geographical coordinates, current temperature and precipitation levels, projected temperature and precipitation changes. This dataset provides a solid basis for a quantitative assessment of the impacts of climate change on crop production and will facilitate the rapidly developing data-driven machine learning applications.

Background & Summary

Climate change affects many processes of food systems directly and indirectly[1], but the primary effects often appear in crop production. Projections of crop production under future climate change have been studied since the early 1980s. From the 1990s onward, researchers have used future climate data and crop simulation models to project the impacts of climate change on crop yields under various scenarios[2]. Since then, crop simulation models have been used in hundreds of studies to simulate yields for different crops under a range of climate scenarios and growing conditions[3]. The results have been periodically reviewed and assessed by national and international organisations, in particular by the Intergovernmental Panel on Climate Change (IPCC) Working Group II, which provides policy-relevant scientific evidence for the impacts of and adaptation to climate change[3]. Review studies covering the last five IPCC assessment cycles confirm that the overall effects are negative but vary significantly among regions[4,5].

Before 2010, simulation studies were conducted mainly by individual research groups using different climate models, target years, spatial resolution with local management and cultivar conditions. Since 2010, however, significant efforts have been made to coordinate modelling studies through Agricultural Model Intercomparison and Improvement Project (AgMIP)[6], which compares results from multiple crop models using standardised inputs. Early AgMIP activities have disentangled sources of uncertainties in crop yield projections and revealed that yield projections are variable among crop models and that model ensemble mean or median often works better than a single model[7–10], underpinning the importance of datasets based on multiple crop models.

Data sets including crop model simulations produced by AgMIP were subjected to statistical analysis and the results were used to quantify the impacts of climate change on major crops[11,12]. A versatile tool to aggregate

[1]Institute for Agro-Environmental Sciences, National Agricultural and Food Research Organization, Tsukuba, Ibaraki, 305-8604, Japan. [2]Institute of Environment and sustainable Development in Agriculture, Chinese Academy of Agricultural Sciences (IEDA,CAAS), Beijing, 100081, China. [3]Alliance of Bioversity International and International Center for Tropical Agriculture (CIAT), Nairobi, Kenya. [4]University of Illinois, Urbana, IL, USA. [5]The University of the West Indies, St. Augustine, Trinidad. [6]IRI THESys, Humboldt-Universität zu Berlin, Berlin, 10099, Germany. [7]Pontificia Universidad Católica de Chile, Santiago, Chile. [8]Applied mathematics and computer science (MIA 518), INRAE AgroParisTech, Université Paris-Saclay, 75231, Paris, France. ✉e-mail: thase@affrc.go.jp
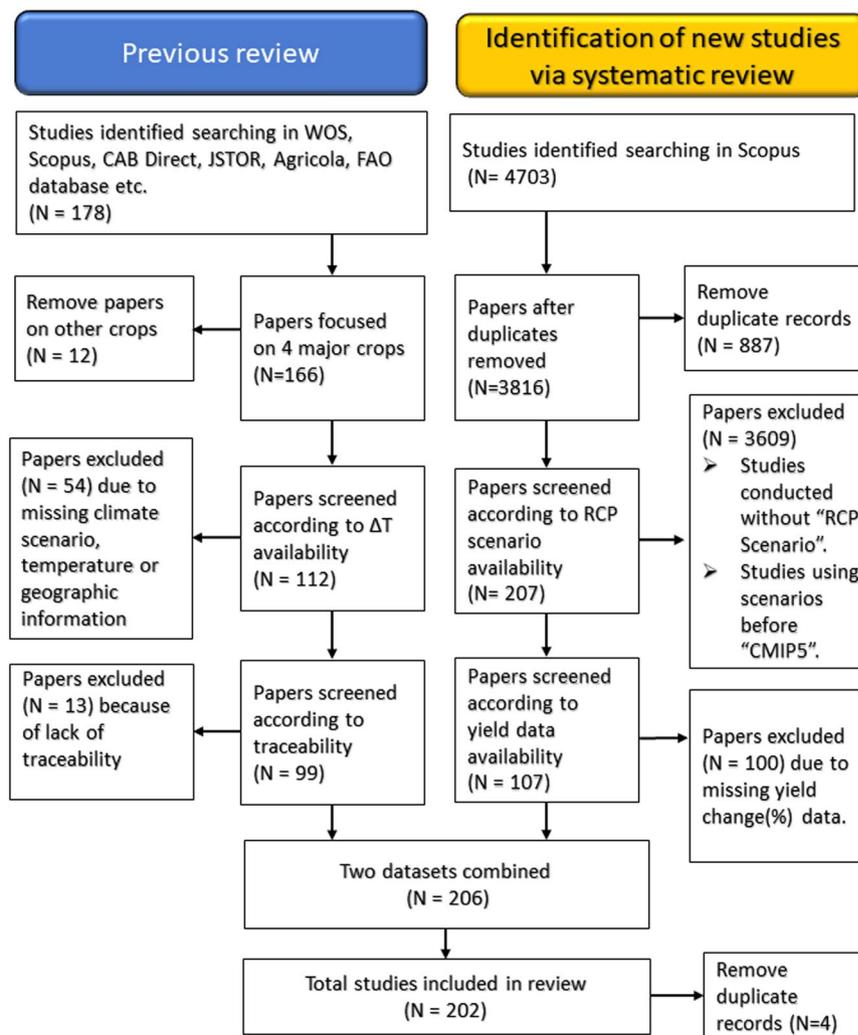
Fig. 1 A diagram depicting paper collection and selection using the two search strategies. N is the number of studies.

simulated results is already available for global gridded studies[13] to facilitate access to the data. Besides these coordinated efforts, however, many simulation results are scattered and not readily available for meta-analysis. To deliver policy-relevant quantitative information, we need to develop a shared and well-documented database that can be used to assess the impacts of different climate and adaptation scenarios on crop yields.

Here, we have developed a global database for potential use for the IPCC Working Group II assessment, obtained through two methods. The first method draws on the dataset used in the meta-analysis of Aggarwal, *et al.*[5], which includes studies considered in the previous five cycles of IPCC assessments[4,14]. The second method is based on a new literature search of studies published during the sixth IPCC assessment cycle (covering the period 2014–2020) reporting crop simulations produced for several contrasting climate change scenarios. The combined dataset covers all six cycles of the IPCC assessment and can serve as a solid basis for analyses from the sixth IPCC assessment onward.

The dataset contains the most relevant variables for evaluating climate change impacts on yields of maize, rice, soybean, and rice for the 21st century. They include geographical coordinates, crop species, CO_2 emission scenarios, CO_2 concentrations, current temperature and precipitation levels, local and global warming degrees, projected changes in precipitation, the relative changes in yield as a percentage of the baseline period obtained with or without CO_2 effects, and with or without implementation of different types of adaptation options.

Methods

Data collection. As shown in a PRISMA diagram (Fig. 1), we obtained data through two methods to develop this dataset. The first method is based on the previous meta-analysis by Aggarwal *et al.*[5], which includes studies published before 2016 (Aggarwal-DS, hereafter). This meta-analysis builds on the dataset used for the 5th IPCC assessment report[4,14] and an additional search through three types of databases: Scientific database (Scopus, Web of Science, CAB Direct, JSTOR, Agricola etc), journals and open access repositories, and institutional Websites (FAO database, AgMIP Database, World Bank, etc.) and Google Scholar. See Aggarwal *et al.*[5] for details. Briefly, the search terms used by Aggarwal *et al.*[5] include "agriculture" or "crop "or "farm" or "crop yield" or "crop yields"
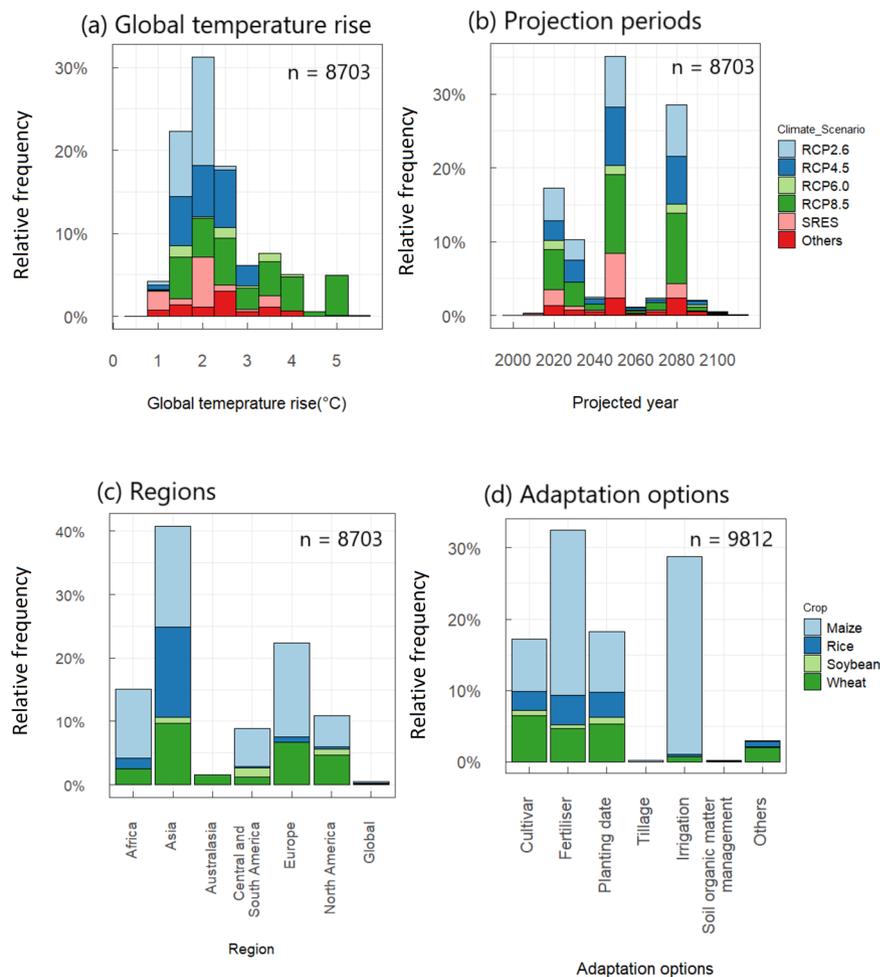
Fig. 2 Data availability of crop yield simulations and its breakdown. **(a)** By global temperature rise from the preindustrial era and climate scenarios, **(b)** By projected time periods (midpoint years) and climate scenarios, **(c)** IPCC regions[29] and crop species, and **(d)** adaptation options and crop species. Note that n = 9812 in adaptation options **(d)** exceeds the total number of simulations (8703) because we collectively add all the options used in the simulations, including those that use multiple options. n is the number of simulation results.

or "farm yields" or "crop productivity" or "agricultural productivity" or "maize" or "rice" or "wheat" and "climate change assessment" or "climate impacts" or "impact assessments" or "climate change impact" or "climate impact" or "effect of climate" or "impact of climate change". The number of selected papers covering the four major crops is 166. We further screened them according to the availability of local temperature rise and geographical information, and traceability, resulting in 99 studies published between 1984 and 2016.

The second method relies on a new recent literature review conducted using Scopus in March 2020 for four major crops (maize, rice, soybean, and wheat) for peer-review papers published from 2014 onward in line with the sixth assessment cycle of IPCC. In this method, we used several combinations of terms to retrieve relevant studies reporting simulations of the impacts of climate change on crop yields using recent climate change scenarios.

For maize, the following search equation was used: PUBYEAR > 2013 AND TITLE-ABS-KEY((maize OR corn) AND (("greenhouse gas" OR "global warming" OR "climate change" OR "climate variability" OR "climate warming")) AND NOT (emissions OR mitigation OR REDD OR MRV)).

Similar search equations were used for the other crops. Collectively, this search returned a total of 4703 references between 2014 and 2020: 1899 for maize, 1790 for wheat, 757 for rice, and 257 for soybean with some duplications because some papers studied multiple crops. Removing the duplicates, the number is down to 3816 studies.

To collect climate-scenario-based simulations, we then selected a subset of studies including the following terms related to climate scenarios in titles, abstracts, or authors' keywords; "RCP", "RCP2.6", "RCP6.0", "RCP4.5", "RCP8.5", "CMIP5", and "CMIP6". RCP stands for the Representative Concentration Pathways[15], and each RCP corresponds to a greenhouse gas concentration trajectory describing different future greenhouse gas emission levels. The number followed by RCP is the level of radiative forcing (Wm^{-2}) reached at the end of the 21st century, which increases with the volume of greenhouse gas emitted to the atmosphere[16]. CMIP5[17] and CMIP6[18]

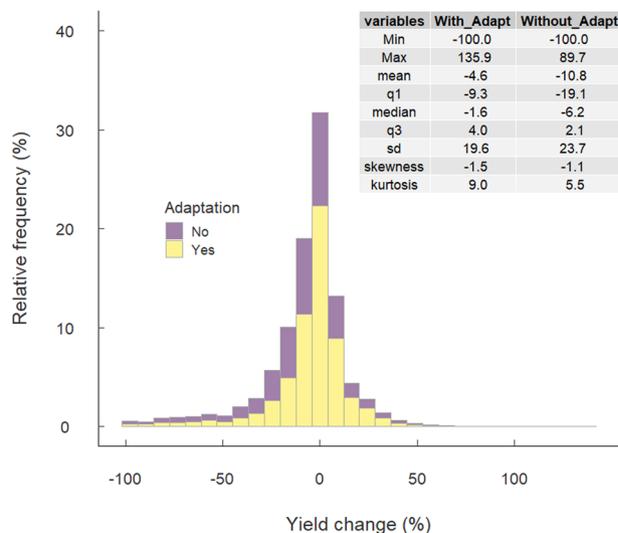| variables	With_Adapt	Without_Adapt
Min	-100.0	-100.0
Max	135.9	89.7
mean	-4.6	-10.8
q1	-9.3	-19.1
median	-1.6	-6.2
q3	4.0	2.1
sd	19.6	23.7
skewness	-1.5	-1.1
kurtosis	9.0	5.5

Fig. 3 Distribution of relative yield change due to climate change from the baseline period (2001–2010) with and without adaptation.

are the Coupled Model Intercomparison Project Phase 5 and Phase 6, respectively, where groups of different earth system models (ESMs) provide global-scale climate projections based on different RCPs. Additionally, "process-based model" was used to search in the authors' keywords to select for studies that use crop simulation models under CMIP5 or CMIP6 climate scenarios. As of March 2020, no results were found for CMIP6 in any search results.

This screening process resulted in a total of 207 references all together for four major crops. These studies were further evaluated for their variables and data availability; studies not reporting yield data were excluded. Projected yields with and without adaptations and yields of the baseline period were extracted, along with geographical coordinates, crop species, greenhouse gas emission scenarios, and adaptation options. We also tried to obtain local and global temperature changes and CO_2 concentrations as much as possible. In addition to extracting data from the literature, we contacted several authors of grid simulation studies to provide aggregated results for countries or regions. The authors of the three grid simulation studies responded and provided baseline and projected yields, annual temperature and precipitation data aggregated over for countries or regions[19–21]. The results from different ESMs were then averaged.

We removed duplicates between the datasets produced by the two methods and ultimately obtained a total of 202 unique studies. Both datasets include studies with different spatial scales: site-based, regional, and global. Among these, the results from the global gridded crop models were aggregated to country levels, and we focused on top-producing countries, which account for 95% of the world's production of each commodity as of 2010 (FAOSTAT, http://www.fao.org/faostat/en/, accessed on September 4, 2020). As a result, the dataset contains 8,703 sets of yield projections during the 21st century from studies published between 1984 and 2020 (Online-only Table 1).

Relative yield impacts. Simulated grain mass per unit land area is used to derive the impact of climate change on yield (YI), which is defined as:

$$YI(\%) = (Y_f/Y_b - 1) \times 100$$

Where Y_f is the future yield, and Y_b is the baseline yield. One study[20] simulated yields separately under both climate change and counterfactual non-climate change scenarios from the pre-industrial era toward the end of the 21st century, also accounting for yield increases due to non-climatic technological factors over time. In this case, YI obtained with the above equation under the climate change scenario was not fully relevant because it combines effects of both climate change and technological factors. Thus, for this study, YI was derived from the average yield in the 2001–2010 period under climate change and the average yield in the same period assuming no climate change, as follows:

$$YI(\%) = [\{(Y_{f\_cc} - Y_{b\_cc}) - (Y_{f\_ncc} - Y_{b\_ncc})\}/Y_{b\_cc} - 1] \times 100$$

Where Y_{f_cc} and Y_{b_cc} are the future and baseline average yields with climate change, Y_{f_ncc} and Y_{b_ncc} are the future and baseline average yields under counterfactual no climate change scenario.

Projected absolute grain yield (t/ha) is also included in the dataset, when available. These values should be used with caution because absolute grains yields are not always comparable due to the use of different yield definitions or assumptions. Different definitions include graded or non-graded yields, husked or unhusked, milled or non-milled yield. Moisture content correction factors can also be different, but these are not often explicitly
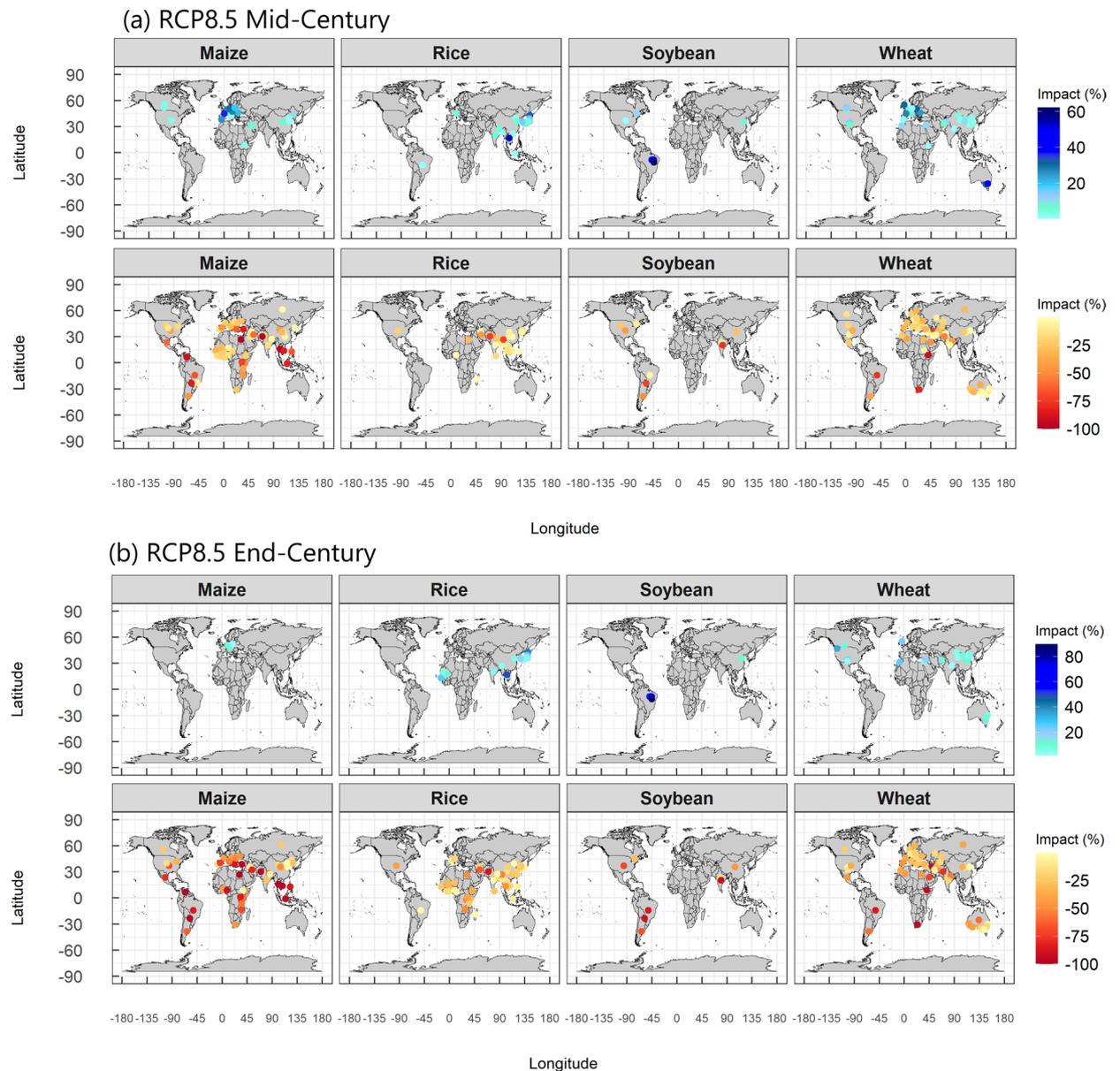
Fig. 4 Climate change impacts (% of yield change from the baseline period) on four crops without adaptation under RCP8.5. (**a**) Mid-century; (**b**) End-Century. Maps with bluish symbols show positive effects (yield gain); Maps with reddish symbols show negative effects (yield loss). Projections under RCP2.6 and RCP4.5 are given in Supplementary Fig. S3.

indicated in the literature. Contrary to absolute yields, relative yields are unitless and rule out differences of yield definitions between studies.

Adaptation to climate changes. Various management or cultivar options are tested in the simulations. If the authors of the article consider these options as ways to adapt crops to climate change, we treat them as adaptation options, which are categorised into fertiliser, irrigation, cultivar, soil organic matter management, planting time, tillage, and others. Specifically, in the fertiliser option, if the amount and timing of fertiliser application are changed from the current conventional method, we treat them as adaptation. In the irrigation option, if the simulation program determines the irrigation scheduling based on the crop growth, climatic and soil moisture conditions, we treat this as adaptation because the management is adjusted to future climatic conditions. If rain-fed and irrigated conditions are simulated separately, we do not consider irrigation as an adaptation. We define cultivar option as the use of cultivars of different maturity groups and/or higher heat tolerance than conventional cultivars. The planting time option corresponds to a shift of planting time from conventional timing. If multiple planting times are tested, we select the one that gives the best yield. The soil organic matter management option corresponds to application of compost and/or crop residue. The tillage option corresponds to reduced- or no-till cultivation compared to no conventional tillage. When studies consider adaptation options, we compute YI from

	Per decade impact (% decade^{-1})				Per degree impact (% °C^{-1})			
	Maize	Rice	Soybean	Wheat	Maize	Rice	Soybean	Wheat
Minimum	−40.0	−40.8	−30.0	−35.4	−158.7	−71.7	−112.6	−122.3
Maximum	14.2	26.2	13.8	21.2	70.8	120.7	58.3	153.7
Mean	−3.9	−1.4	−2.6	−1.8	−13.5	−2.6	−8.8	−5.6
1st quartile	−5.5	−2.5	−6.7	−3.5	−18.1	−7.1	−16.9	−10.9
Median	−2.1	−0.7	−1.2	−1.2	−7.1	−2.3	−4.0	−3.7
3rd quartile	−0.3	0.8	1.7	0.7	−1.1	2.7	6.3	2.3
Standard deviation	7.0	4.7	7.4	5.0	25.4	12.0	26.3	17.4
Skewness	−1.8	−2.6	−0.9	−1.7	−1.9	1.0	−1.1	−1.2
Kurtosis	8.6	21.9	4.7	10.8	8.6	24.4	5.2	17.4

Table 1. Summary statistics of climate change impacts on four major crops expresses as per decade impact and per degree impact without adaptation.

the ratio of yield with adaptation under climate change to baseline yield without adaptation. To measure our capacity to adapt to climate change, we calculated adaptation potential - defined as the difference between yield impacts with and without adaptation - when a pair of yield values were available in the same study.

Temperature and precipitation changes. Both local temperature rise (ΔT_l) and global mean temperature rise (ΔT_g) from the baseline period have important implications. The former directly affects crop growth and yield, and the latter represents a global target associated with mitigation activities. We extracted both ΔT_l and ΔT_g from the literature as much as possible, but ΔT_g is not available in many studies. In such cases, we estimated ΔT_g using the Warming Attribution Calculator (http://wlcalc.climateanalytics.org/choices). In the dataset, we provide two estimates for ΔT_g: one from the current baseline period (2001–2010) and the other from the preindustrial era (1850–1900). We also extracted precipitation changes (ΔPr) and baseline precipitation data reported in the selected studies. When only relative changes were available for precipitation data, we estimated ΔPr using the reported relative change and current precipitation levels described in the next section.

Current temperature and precipitation levels. Current annual mean temperatures and precipitation were obtained from the W5E5 dataset[22], which was compiled to support the bias adjustment of climate input data for the impact assessments performed in Phase 3b of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP3b, https://www.isimip.org/protocol/3/). The W5E5 dataset includes half-degree grid resolution daily mean temperature and precipitation data from 1979 to 2016, which we averaged for the period from 2001 and 2010. They were then extracted for each simulation site or region using the geographic information. For global simulations, which were aggregated to the country level, central coordinates were used to link gridded temperature and precipitation data with each country. As centroids may not represent the centre of the growing regions, particularly in large countries, growing-area weighted averages of annual temperature and precipitation were also provided using MIRCA 2000[23], which contains half-degree grid harvested areas (a total of irrigated and rainfed) around the year 2000.

CO$_2$ concentrations. Several studies report two series of yield simulations obtained using two CO$_2$ levels to infer the CO$_2$ fertilization effects: one obtained with CO$_2$ concentrations fixed at the current levels and the other obtained with increased future CO$_2$ concentrations provided by the emission scenario considered. In the dataset, we make this explicit in the following two variables:

1. CO$_2$: Binary variable equal to "Yes" if future CO$_2$ concentrations from the emission scenarios were used and "No" if the current CO$_2$ concentration was used for the yield simulations.
2. CO$_2$ ppm: if available, CO$_2$ concentration was extracted from the original paper. If not, we calculated it from projected changes in CO$_2$ concentrations based on the scenarios and periods studied. CO$_2$ concentration data were obtained from https://www.ipcc-data.org/observ/ddc_co2.html for CMIP3 and Meinshausen, et al.[16] (http://www.pik-potsdam.de/~mmalte/rcps/) for CMIP5.

Baseline correction. Because baseline periods differed between studies, we corrected YI, ΔT_l, ΔT_g, ΔPr to the 2001–2010 baseline period by a linear interpolation method following Aggarwal et al.[5]. Namely, the impacts YI were first divided by the year gap between the future period midpoint year and the baseline period midpoint year of the original study. The impact per year was then multiplied by the year gap from our reference baseline period midpoint year (2005). The same method was applied to express ΔT_l and ΔPr relatively to 2001–2010.

We made all data publicly available to increase accessibility (see Data Records section for access).

Data Records

All the data and R scripts associated with the dataset are stored in the figshare repository[24], where the following files are uploaded:
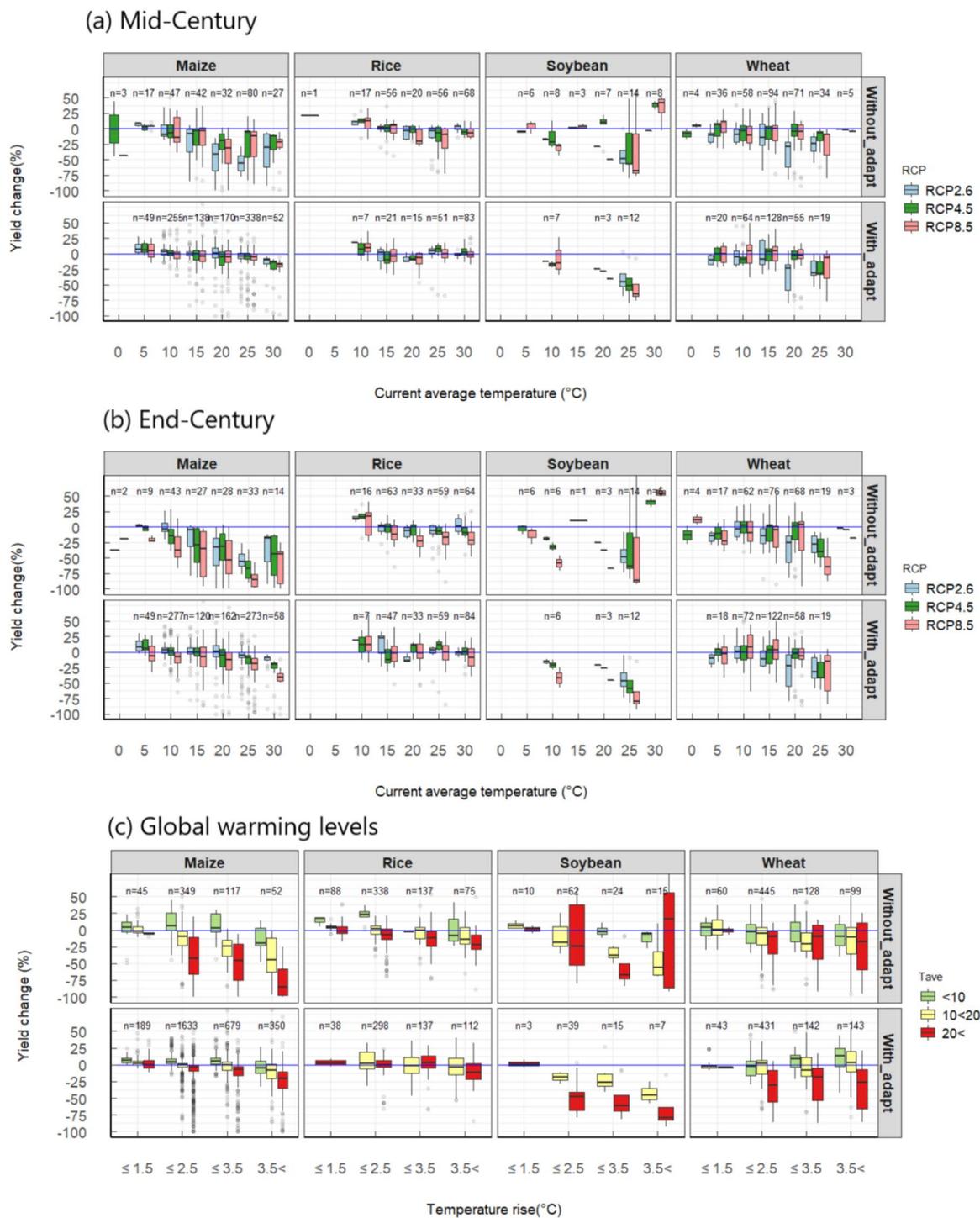
Fig. 5 Projected yield changes relative to the baseline period (2001–2010). **(a)** Mid-century (MC) projections without adaptation under RCP8.5 scenario, upper panels showing positive impacts and lower panels negative impacts, **(b)** End-century (EC) projections under three RCP scenarios by current annual temperature (T_{ave}), and **(c)** Yield change as a function of global temperature rise from the pre-industrial period by three T_{ave} levels. The box is the interquartile range (IQR) and the middle line in the box represents the median. The upper- and lower-end of whiskers are median $1.5 \times IQR \pm$ median. Open circles are values outside the $1.5 \times IQR$.

1. "Projected_impacts_datasheet_11.24.2021.xlsx" includes three worksheets. "Projected_impacts" worksheet contains the final dataset after screening, and "Adaptation_potential" is the extracted subset of the paired data comparing yield impacts with and without adaptation. "Excluded" has untraceable simulation results in the Aggarwal-DS.
2. "Meta-data_11.25.2021.xlsx" contains the summary of the dataset, such as the definition and unit of the variables used in "Projected_impacts_datatasheet.xlsx".
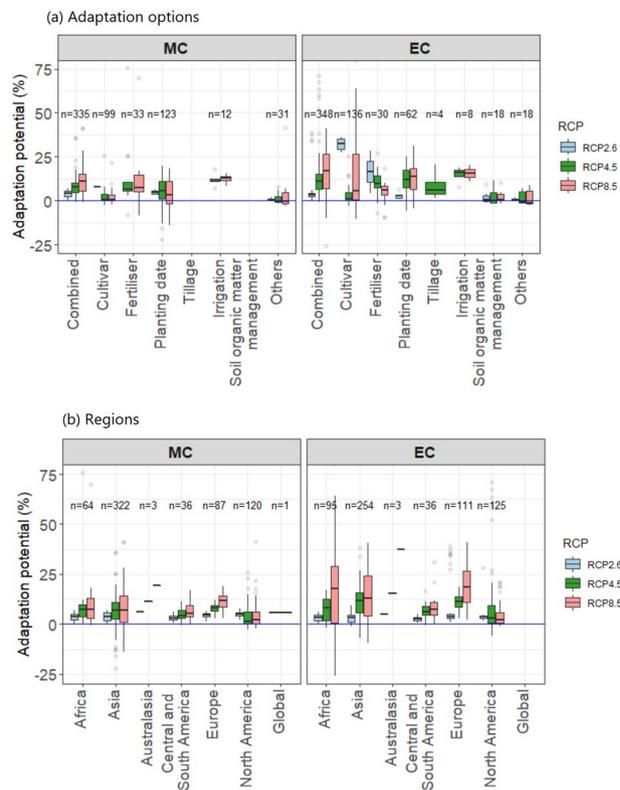
Fig. 6 Adaptation potential, defined as the difference between yield impacts with and without adaptation in projected yield impacts, for three RCPs by mid- and end-century (MC, EC). The box is the interquartile range (IQR) and the middle line in the box represents the median. The upper- and lower-end of whiskers are median $1.5 \times IQR \pm$ median. Open circles are values outside the $1.5 \times IQR$. **(a)** By adaptation options and **(b)** by IPCC regions.

. . . .

3. "Online_only_summary_tables_11.18.2021.xlsx" contains data distribution, median, and mean impacts of climate change, presented in the online-only tables.
4. "Supplementary_materials_11.29.2021.pdf" contains methods for estimating local temperature rise and summary distribution of climate change impacts on four crop yields.
5. "Reference_11.24.2021.docx" provides a list of references that provided data.
6. "R_script_for_Hasegawa_et.al.11.26.2021.zip" contains R scripts used to estimate missing values of $\Delta T_h \Delta T_l$ and ΔPr and draw Figs. 2–6.

Coverage of the data. A total of 8703 yield simulations are registered in the consolidated dataset. The number of simulations grows exponentially with publication year: 20 in the 1980s, 304 in 1990s, 830 in 2000s and 7549 in 2010s (Online-only Table 1). About 80% of the simulations use CMIP5 climate scenarios, and 11% use CMIP3. From CMIP5, RCP2.6, RCP4.5 and RCP8.5 are the most used concentration pathways (Online-only Table 2, Fig. 2a). ΔT_g from the baseline period (2001–2010) ranges from 0 to 4.8 °C (0.8 to 5.6 °C from the preindustrial period). Almost all simulations with $\Delta T_g > 3$ °C use RCP8.5, resulting in a greater ΔTg range under CMIP5 (RCPs) than under previous scenarios (SRES and others).

Projected time periods span widely in the 21st century, but the midpoint years peak at 2020 for the near future, 2050 for mid-century, and 2080 for end-century (Fig. 2b). Major emission scenarios such as RCP2.6, 4.5 and 8.5 are almost equally distributed across time periods. About 5% of the simulations assume no CO_2 fertilisation effects.

Relative frequency of the regions studied generally reflects harvested areas of the four crops in each region (Fig. 2c). About 41% of the simulations were performed in Asia, which accounts for about 47% of the harvested area of the four major crops (mean of 2017–2019, FAOSTAT, http://www.fao.org/faostat/en/, accessed on April 28, 2021). Europe is slightly overstudied (22%) for its world share of the harvested areas (12%). Central and South Americas is slightly under-researched (9%) for the regional share of harvested areas (15%), whereas Africa's share (15%) is comparable to the area harvested (10%). Altogether global harvested areas for these four major crops is 7×10^8 ha: wheat represents 31% of this area, followed by maize (28%), rice (23%) and soybean (18%). Maize studies are over represented, accounting for about half of the simulations (52%), followed by wheat (26%) and rice (17%); soybean accounts only for 3% of the simulations (Fig. 2c). Regionally, maize and wheat are harvested across almost all regions, and simulations follow the actual distribution of these crops. Rice is predominantly studied in Asia, reflecting actual distribution (85% of the harvested area is in Asia). Soybean

remains understudied compared to the other three crops despite its large cultivated area (about 75% of the rice harvested area). Regionally, simulation sites or regions for soybean are mostly in the Americas, which account for 76% of the total soybean harvested area.

About 39% of the simulations (3376) use current management practices, and the rest (5327) consider different management or cultivars as adaptation options (Fig. 1d). More than half of the simulations are run with multiple options. Among these options, fertiliser accounts for 32% followed by irrigation (29%), cultivar and planting date (17% each). There are 2005 pairs of yield simulations available for comparing results obtained with and without adaptation. These pairs of yield data can be used to compute the adaptation potentials of the different options considered.

Technical Validation

Data quality check. We repeatedly checked the data with multiple authors for the new dataset. For the Aggarwal-DS, we reviewed the sources of references. In case of missing information such as climate scenarios, CO_2 concentration, or temperature increase, we came back to the original reference. Inconsistencies between the dataset and original papers were corrected when possible. Overall, corrections were made on 333 simulations from 10 studies, which we flag with "*" in the remark column of the dataset. We removed all data of the Aggarwal-DS that were untraceable in the original paper. This quality control excluded 47 simulations from 9 articles listed in the "Excluded" sheet.

We first examined the distribution of the climate change impacts on crop yields, which span from −100 to 136% (Fig. 3). This distribution is skewed to the left, as indicated by the negative skewness. The large kurtosis shows that distribution tails are longer than than those of the normal distribution. We tested the effects of potential outliers outside the 1.5-fold interquartile range (IQR) on the summary statistics of the climate change impacts on crop yields[25]. Removing values outside the 1.5-fold IQR decreases the number of simulations by 907(10.4%) and the negative effects of climate change on crop yields by 3.0% for the mean and 0.6% for the median, suggesting that the deletion affected the original distribution. We, therefore, keep all the simulation results in the dataset.

Methods to estimate local temperature and precipitation changes. Out of 8703 simulations, local temperature change (ΔT_l) and global temperature range (ΔT_g) were available in 4316 and 8109 simulations, respectively. To estimate ΔT_l for 3793 simulations with missing ΔT_l, we examined the relationship between ΔT_l and the following six input variables in 4316 simulations where ΔT_l was available: ΔT_g, average temperature (area weighted), latitudes, longitudes, time periods, and emission scenarios. Values of ΔT_l were estimated using random forest algorithms trained to establish a function relating local temperature rise to the six inputs considered. We tested and compared four models based on different combinations of the input variables. Among the four models, a reduced model with three variables (ΔT_g, latitude, and longitude) showed the highest percentage of explained variance (97.1%), and led to a cross-validated RMSE as low as 0.18 °C (Supplementary Table S1 and Fig. S1). We, therefore, used the reduced model to impute ΔT_l for the 4430 missing data. We also estimated ΔT_g for 504 simulations with missing ΔT_l from ΔT_g, average temperature (area weighted), latitude, longitude, climate scenarios, future-midpoint year (Supplementary Table S2 and Fig. S2).

Likewise, we applied a random forest model to estimate ΔPr from current annual precipitation and average temperature (area weighted), latitude, longitude, local temperature change from 2005), climate scenario, future mid-point year, and climate change impact on yield relatively to 2005. Among eight models tested, a one with ΔT_g, ΔT_l, latitude, longitude, RCP, future-mid-point year and current annual precipitation perfomed best, which accounted for 96.9% of the out-of-bag variation of the data (n = 3560) and led to a cross-validated RMSE was 18 mm (Supplementary Table S3). We then applied this model to estimate all missing ΔPr.

Comparison with previous studies. The overall effects of climate change on crop yields are negative, with the mean and median of −11% and −6.2% without adaptation and −4.6% and −1.6% with adaptation, respectively (Online-only Tables 3 and 4). The median per-decade yield impact without adaptation is −2.1% for maize, −1.2% for soybean, −0.7% for rice, and −1.2% for wheat (Table 1), which are consistent with previous IPCC assessments[14]. The median per-warming-degree impact is −7.1% for maize, −4.0% for soybean, −2.3% for rice, and −3.7% for wheat (Table 1). Per-degree yield impacts for each crop are generally within the range reported in the previous meta-analysis[11]. Among the four crops, soybean has the least number of simulations, resulting in a greater variation in both per-decade and per-degree impacts. Maize consistently shows the largest negative impacts, while rice shows the least.

The climate change impacts by IPCC regional groups reveals that Europe and North America are expected to be less affected by climate change in the mid-century (MC) and the end-century (EC) than Africa, Central and South America, particularly for maize and soybean. Both positive and negative effects are mixed in all regions (Fig. 4, Supplementary Figs. S3, S4).

Regional differences in the impacts in MC and EC are associated with the current temperature level. In MC, positive or neutral effects are projected when current annual average temperatures (T_{ave}) are below 10–15 °C, but the effects become negative as T_{ave} increases beyond these levels regardless of RCPs (Fig. 5a). This accounts for the regional differences as a function of latitude reported in previous meta-analyses[4,5]. In EC, the threshold T_{ave} shifts lower, and the negative effects become more severe, particularly under a high emission scenario (RCP8.5) (Fig. 5b). The effect of ΔT_g from the baseline period on YI differs depending on the T_{ave} (Fig. 5c); At $T_{ave} < 10$ °C, YI is generally neutral even where $\Delta T_g > 2$ °C in most crops, but at $T_{ave} > 20$ °C, YI is negative even with small ΔT_g, notably in maize. The difference in the YI dependence on ΔT_g between regions is also consistent with the previous study[4].

Adaptation potential averaged 7.3% in MC and 11.6% in EC (Fig. 6, Supplementary Fig. S5), which is not sufficient to offset the negative impacts, particularly in currently warmer regions. Residual damages will thus likely remain even with adaptation, which is also supported by other lines of evidence[26,27].

Usage Notes

Crop yield simulation studies can provide a narrative of when, where, and what will happen to crop production under different GHG emissions and climate scenarios. They are also expected to provide quantitative information on the potential and limits to adaptation. However, robust estimates covering different temporal and spatial scales need to draw on multiple results obtained from various simulation studies. Nearly four decades have passed since the model projections based on future climate scenarios started. This dataset covers the entire period of simulation studies using climate scenarios, which can help update the quantitative review of climate change impacts on crops. The full list of references is provided in the reference file (https://doi.org/10.6084/m9.figshare.14691579.v4).

Currently, studies are heavily biased towards major cereals such as maize, rice, and wheat, but this can be expanded to include other crops. As of 2020, our literature search failed to find published reports using CMIP6 climate scenarios, but this dataset can be easily updated when new simulations using new climate scenarios or other crop species become available. The next IPCC assessment cycle can fully utilise this dataset by adding the latest simulation results.

One of the caveats to the current dataset is that it only includes crop yield data, notwithstanding crop simulation studies are expected to produce other results than yield. Because of the recent progress in crop modelling, grain quality projections are emerging[28]. We have extensively included the temperature and precipitation levels to account for the impacts concerning the warming and current temperature, but there is a need to include other key climatic variables such as soil moisture. It will be useful to expand our dataset in the future to include this type of data.

Code availability

Script files were created using the R statistical programming to estimate missing values of ΔT_l, ΔT_l and ΔPr and draw Figs. 2–6 which are available in the figshare repository[24].

References

1. Davis, K. F., Downs, S. & Gephart, J. A. Towards food supply chain resilience to environmental shocks. *Nature Food* **2**, 54–65, https://doi.org/10.1038/s43016-020-00196-3 (2021).
2. Parry, M. L. *et al.* in *The IPCC Impacts Assessment* (eds W. J. M Tegart, G. W. Sheldon, & D.C. Griffiths) 1–45 (AGPS Press, Canberra, Australia, 1990).
3. Porter, J. R. *et al.* Invited review: Intergovernmental Panel on Climate Change, agriculture, and food—A case of shifting cultivation and history. *Global Change Biology* **25**, 2518–2529, https://doi.org/10.1111/gcb.14700 (2019).
4. Challinor, A. J. *et al.* A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change* **4**, 287–291, https://doi.org/10.1038/nclimate2153 (2014).
5. Aggarwal, P., Vyas, S., Thornton, P. & Campbell, B. M. How much does climate change add to the challenge of feeding the planet this century? *Environmental Research Letters* **14**, https://doi.org/10.1088/1748-9326/aafa3e (2019).
6. Rosenzweig, C. *et al.* The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology* **170**, 166–182, https://doi.org/10.1016/j.agrformet.2012.09.011 (2013).
7. Asseng, S. *et al.* Uncertainty in simulating wheat yields under climate change. *Nature climate change* **3**, 827–832 (2013).
8. Li, T. *et al.* Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. *Global Change Biology* **21**, 1328–1341, https://doi.org/10.1111/gcb.12758 (2015).
9. Bassu, S. *et al.* How do various maize crop models vary in their responses to climate change factors? *Global Change Biology* **20**, 2301–2320, https://doi.org/10.1111/gcb.12520 (2014).
10. Wallach, D. *et al.* Multimodel ensembles improve predictions of crop–environment–management interactions. *Global Change Biology*, 5072–5083, https://doi.org/10.1111/gcb.14411 (2018).
11. Zhao, C. *et al.* Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences* **114**, 9326–9331, https://doi.org/10.1073/pnas.1701762114 (2017).
12. Makowski, D., Marajo-Petitzon, E., Durand, J.-l & Ben-Ari, T. Quantitative synthesis of temperature, CO2, rainfall, and adaptation effects on global crop yields. *European Journal of Agronomy* **115**, 126041–126041, https://doi.org/10.1016/j.eja.2020.126041 (2020).
13. Villoria, N. B. *et al.* Rapid aggregation of global gridded crop model outputs to facilitate cross-disciplinary analysis of climate change impacts in agriculture. *Environmental Modelling & Software* **75**, 193–201, https://doi.org/10.1016/j.envsoft.2015.10.016 (2016).
14. Porter, J. R. *et al.* in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Repor+* (eds C. B. Field *et al.*) 485–533 (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014).
15. van Vuuren, D. P. *et al.* The representative concentration pathways: an overview. *Climatic Change* **109**, 5, https://doi.org/10.1007/s10584-011-0148-z (2011).
16. Meinshausen, M. *et al.* The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change* **109**, 213–241, https://doi.org/10.1007/s10584-011-0156-z (2011).
17. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society* **93**, 485–498, https://doi.org/10.1175/bams-d-11-00094.1 (2012).
18. Eyring, V. *et al.* Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016 (2016).
19. Ishigooka, Y., Hasegawa, T., Kuwagata, T. & Nishimori, M. Evaluation of the most appropriate spatial resolution of input data for assessing the impact of climate change on rice productivity in Japan. *Journal of Agricultural Meteorology* **76**, 61–68, https://doi.org/10.2480/agrmet.D-19-00021 (2020).
20. Iizumi, T. *et al.* Responses of crop yield growth to global temperature and socioeconomic changes. *Scientific Reports* **7**, 7800–7800, https://doi.org/10.1038/s41598-017-08214-4 (2017).

21. Deryng, D., Sacks, W. J., Barford, C. C. & Ramankutty, N. Simulating the effects of climate and agricultural management practices on global crop yield. *Global Biogeochemical Cycles* **25**, https://doi.org/10.1029/2009GB003765 (2011).
22. Lange, S. NWFDE5 over land merged with ERA5 over the ocean (W5E5). V. 1.0. *GFZ Data Services* https://doi.org/10.5880/pik.2019.023 (2019).
23. Portmann, F. T., Siebert, S. & Döll, P. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles* **24**, https://doi.org/10.1029/2008GB003435 (2010).
24. Hasegawa, T. *et al.* A global dataset for the projected impacts of climate change on four major crops. *figshare* https://doi.org/10.6084/m9.figshare.14691579.v4 (2021).
25. Upton, G. & Cook, I. *Understanding Statistics*. (Oxford University Press, 1997).
26. Minoli, S. *et al.* Global Response Patterns of Major Rainfed Crops to Adaptation by Maintaining Current Growing Periods and Irrigation. *Earth's Future* **7**, 1464–1480, https://doi.org/10.1029/2018EF001130 (2019).
27. Iizumi, T. *et al.* Climate change adaptation cost and residual damage to global crop production. *Climate Research* **80**, 203–218, https://doi.org/10.3354/cr01605 (2020).
28. Asseng, S. *et al.* Climate change impact and adaptation for wheat protein. *Global Change Biology* **25**, 155–173, https://doi.org/10.1111/GCB.14481 (2018).
29. Hewitson, B. *et al.* in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* 1133–1197 (Cambridge University Press, Cambrige, UK, 2014).

Author contributions

Toshihiro Hasegawa and Hitomi Wakatsuki designed the dataset. Hitomi Wakatsuki and Hui Ju collected simulation results from the SCOPUS search. Shalika Vyas designed and collected the Aggarwal dataset. Gerald C. Nelson conducted literature search and provided global temperature dataset. David Makowski and Hitomi Wakatsuki developed a statistical imputation for missing data on the local temperature rise and precipitation change. All authors worked on data analysis and drafting the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-022-01150-7.

Correspondence and requests for materials should be addressed to T.H.

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

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