Spatial and temporal uncertainty of crop yield aggregations

Vera Porwollik[a], Christoph Müller[a], Joshua Elliott[b,c], James Chryssanthacopoulos[c], Toshichika Iizumi[d], Deepak K. Ray[e], Alex C. Ruane[c,f], Almut Arneth[g], Juraj Balkovič[h,i], Philippe Ciais[j], Delphine Deryng[b,c], Christian Folberth[h,k], Roberto C. Izaurralde[l,m], Curtis D. Jones[l], Nikolay Khabarov[h], Peter J. Lawrence[n], Wenfeng Liu[o], Thomas A.M. Pugh[g,p], Ashwan Reddy[l], Gen Sakurai[d], Erwin Schmid[q], Xuhui Wang[j,r], Allard de Wit[s], Xiuchen Wu[j]

Corresponding author

Tel.: +49 331 288 20824; email: vera.porwollik@pik-potsdam.de; Potsdam Institute for Climate Impact Research, Telegraphenberg 31, 14473 Potsdam, Germany

Highlights

- We aggregate 14 simulated gridded crop yields with four harvested areas data sets
- Uncertainties in multi-annual means and temporal patterns are quantified
- Aggregation uncertainties can be substantial but are often small
- Aggregation uncertainty should be considered in model evaluation and impact studies

Keywords
Aggregation uncertainty, global crop model, crop yields, gridded data, harvested area

[a] Potsdam Institute for Climate Impact Research, Research Domain II Climate Impacts & Vulnerabilities, 14473 Potsdam, Germany
[b] Columbia University Center for Climate Systems Research, NASA Goddard Institute for Space Studies, New York, NY 10025, USA
[c] University of Chicago and ANL Computation Institute, Chicago, IL 60637, USA
[d] National Agriculture and Research Organization, Institute for Agro-Environmental Sciences, Tsukuba, 305-8604, Japan
[e] University of Minnesota, Institute on the Environment, Saint Paul, MN 55108, USA
[f] NASA Goddard Institute for Space Studies, New York, NY 10025, USA
[g] Karlsruhe Institute of Technology, IMK-IFU, 82467 Garmisch-Partenkirchen, Germany
[h] International Institute for Applied Systems Analysis, Ecosystem Services and Management Program, 2361 Laxenburg, Austria
[i] Comenius University in Bratislava, Department of Soil Science, 842 15 Bratislava, Slovak Republic
[j] Laboratoire des Sciences du Climat et de l'Environnement, CEA CNRS UVSQ, Orme des Merisiers, 91191 Gif-sur-Yvette, France
[k] Ludwig Maximilian University, Department of Geography, 80333 Munich, Germany
[l] University of Maryland, Department of Geographical Sciences, College Park, MD 20742, USA
[m] Texas A&M University, Texas AgriLife Research and Extension, Temple, TX 76502, USA
[n] National Center for Atmospheric Research, Earth System Laboratory, Boulder, CO 80307, USA
[o] Swiss Federal Institute of Aquatic Science and Technology, Eawag, CH-8600 Duebendorf, Switzerland
[p] University of Birmingham, School of Geography, Earth & Environmental Science and Birmingham Institute of Forest Research, B15 2TT Birmingham, United Kingdom
[q] University of Natural Resources and Life Sciences, Institute for Sustainable Economic Development, 1180 Vienna, Austria
[r] Peking University, Sino-French Institute of Earth System Sciences, 100871 Beijing, China
[s] Alterra Wageningen University and Research Centre, Earth Observation and Environmental Informatics, 6708PB Wageningen, Netherlands

Abstract

The aggregation of simulated gridded crop yields to national or regional scale requires information on temporal and spatial patterns of crop-specific harvested areas. This analysis estimates the uncertainty of simulated gridded yield time series related to the aggregation with four different harvested area data sets. We compare aggregated yield time series from the Global Gridded Crop Model Intercomparison project for four crop types from 14 models at global, national, and regional scale to determine aggregation-driven differences in mean yields and temporal patterns as measures of uncertainty.

The quantity and spatial patterns of harvested areas differ for individual crops among the four data sets applied for the aggregation. Also simulated spatial yield patterns differ among the 14 models. These differences in harvested areas and simulated yield patterns lead to differences in aggregated productivity estimates, both in mean yield and in the temporal dynamics.

Among the four investigated crops, wheat yield (17% relative difference) is most affected by the uncertainty introduced by the aggregation at the global scale. The correlation of temporal patterns of global aggregated yield time series can be as low as for soybean (r=0.28).

For the majority of countries, mean relative differences of nationally aggregated yields account for 10% or less. The spatial and temporal difference can be substantial higher for individual countries. Of the top-10 crop producers, aggregated national multi-annual mean relative difference of yields can be up to 67% (maize, South Africa), 43% (wheat, Pakistan), 51% (rice, Japan), and 427% (soybean, Bolivia). Correlations of differently aggregated yield time series can be as low as r=0.56 (maize, India), r=0.05 (wheat, Russia), r=0.13 (rice, Vietnam), and r=-0.01 (soybean, Uruguay). The aggregation to sub-national scale in comparison to country scale shows that spatial uncertainties can cancel out in countries with large harvested areas per crop type. We conclude that the aggregation uncertainty can be substantial for crop productivity and production estimations in the context of food security, impact assessment, and model evaluation exercises.

1. Introduction

Crop models are increasingly applied at the global scale to study how agricultural yields and total production over regions might be affected by global phenomena such as market dynamics and climate change. Simulations of crop productivity (yield) at different spatial and temporal scales have been used for example in the context of food security, land use, and climate change (Asseng et al., 2015; Challinor et al., 2014; Mueller et al., 2012; Nelson et al., 2014a,b). Uncertainties associated with crop model projections have been widely recognized and discussed, including those attributed to input uncertainty (Roux et al., 2014), as to differences in climate forcing data (Rosenzweig et al., 2014), model structure and parameterization (Rötter et al., 2012), and assumptions on the effectiveness of CO_2-fertilization on crop yields (Deryng et al., 2014). The uncertainty in cropland extent and its implications for land use modeling have been addressed before by Eitelberg et al. (2015), Fritz et al. (2015), and See et al. (2015).

Gridded cropping system data sets on the spatial distribution of crops at the global scale have been reported by Leff et al. (2004), and more recently by Iizumi et al. (2014), and Ray et al. (2012) including distinct data on crop-specific harvested area. Anderson et al. (2015) directly compared four gridded cropping system data sets as MIRCA2000 (Portmann et al., 2010), SPAM2000 (You et al., 2014), GAEZ (Fischer et al., 2012), and M3 (Monfreda et al., 2008). They conclude that the data sets' differences in harvested area and yield could be attributed mainly to the input data used and the downscaling method applied, and report that the disagreement between data sets was largest in areas with minimal harvested area. Different schemes for the interpolation of site-specific yields for the aggregation to agro-climatic zones have been discussed by van Wart et al. (2013) within the context of yield gap and production analysis.

Global gridded crop model (GGCM) results e.g. yield (t/ha) are typically reported in a standardized half degree grid format. This output is aggregated at annual time steps to different spatial scales within the context of model skill assessment, impact studies, or as input variable to land use models. It is used for example when comparing different countries or evaluating modeled yields against

agricultural statistics that are only available at the aggregated scale of administrative units. For this kind of aggregation, data sets on spatial patterns of crop-specific harvested area are applied, which are typically derived from data on cropland extent, national and sub-national census data, and allocation rules. To date, little attention has been paid to the uncertainty of aggregation of gridded crop model simulations induced by the choice of crop-specific harvested area data set. Thus the objective of this study is to assess this aggregation uncertainty at different spatial scales. We use the term "crop mask" in the following as a short version of "gridded crop-specific harvested area data set". The uncertainty in simulated yields related to aggregation masks is determined by two factors: a) the differences in quantity and spatial patterns of crop-specific harvested area data sets, and b) the spatial and quantitative heterogeneity of simulated crop yields, which is specific to individual GGCMs.

2. Material and methods

2.1 Model input data and crop yield simulations

In the Global Gridded Crop Model Intercomparison (GGCMI) project Phase 1 (http://www.agmip.org/ag-grid/ggcmi/) of the Agricultural Model Intercomparison and Improvement Project (AgMIP) (Rosenzweig et al., 2013) 14 modeling groups performed historical global crop growth simulations according to the modeling protocol of Elliott et al. (2015). Crop growth has been simulated using the bias-corrected historical weather input data sets AgMERRA (Ruane et al., 2015) and the atmospheric CO_2-data based on the Mauna Loa Observatory time series (Thoning et al., 1989). AgMERRA provides daily data for the time period 1980-2010 and had been aggregated from the original resolution of 0.25° to 0.5° before being supplied to modelers. The Mauna Loa Observatory time series reports observed annual and monthly values of the atmospheric CO_2-mixing ratio, so that models simulated crop growth with a CO_2-mixing ratio of 339-390ppmv (here stating annual averages 1980-2010).

Four crop types were simulated by the modeling teams: maize (Zea mays L.), wheat (*Triticum aestivum* L.), rice (*Oryza sativa* L.), and soybean (*Glycine max* (L.) Merr.) These crops had been

categorized in the GGCMI project as Priority 1 crops, because of their importance as agricultural commodity in terms of their global harvested area covered, production amount, level of trade, and direct or indirect contribution to human diet.

The participating models cover a broad range of model types and of implemented processes. Their basic characteristics and key literature references are listed in Table 1 (more details in *SI Appendix* Tables A.1-5).

Table 1: Participating models in the study

Crop model	Model type	Key literature
CGMS-WOFOST	Empirical/process hybrid	de Wit van Diepen (2008)
CLM-Crop	Dynamic Global Vegetation Model	Drewniak et al. (2013)
EPIC-BOKU	Site-based process model (based on EPIC)	EPIC v0810 - Izaurralde et al. (2006), Williams (1995)
EPIC-IIASA	Site-based process model (based on EPIC)	Izaurralde et al. (2006), Williams (1995)
EPIC-TAMU	Site-based process model (based on EPIC)	EPIC v1102- Izaurralde et al. (2012)
GEPIC	Site-based process model (based on EPIC)	EPIC v0810 - Liu et al. (2007), Williams (1995)
LPJ-GUESS	Dynamic Global Vegetation Model	Lindeskog et al. (2013), Smith et al. (2001)
LPJmL	Dynamic Global Vegetation Model	Waha et al. (2012), Bondeau et al. (2007)
ORCHIDEE-crop	Dynamic Global Vegetation Model	Wu et al. (2015)
pAPSIM	Site-based process model	APSIM v7.5 - Elliott et al. (2014), Keating et al. (2003)
pDSSAT	Site-based process model	pDSSAT v1.0 - Elliott et al. (2014); DSSAT v4.5 - Jones et al. (2003)
PEGASUS	Empirical/process hybrid	v1.1- Deryng et al. (2014), v1.0 - Deryng et al. (2011)
PEPIC	Site-based process model (based on EPIC)	EPIC v0810- Liu et al. (2016), Williams (1995)
PRYSBI2	Empirical/process hybrid	Sakurai et al. (2014)

For the crop growth simulations initial conditions of soil water, minerals, crop residues, and soil organic matter were derived by applying different soil input data and spin-up runs individual to each of the modeling groups (*SI Appendix* Table A.3). Modelers were asked to model all crops wherever a given crop can grow and at least on all current agricultural land. The GGCMI project distinguishes three levels of model harmonization with respect to agricultural management. We here used the simulations of the “default” model configuration if available, where every modeling team used their own assumptions on agricultural management (varieties, growing season, fertilizer etc.). The EPIC-TAMU model was run at the global scale for the first time and ORCHIDEE-crop never globally simulated soybean before and thus could not provide a “default” simulation. These teams used the global input data on sowing and maturity dates, and fertilizer data provided within the context of the

GGCMI project for a rather harmonized simulation, so that for this study their "fullharm" model configuration was used. The modeling teams reported two separate yield time series per configuration type - one assuming rainfed and the other fully irrigated production conditions everywhere. The irrigated crop growth simulations were run assuming unlimited water supply without conveyance or application losses.

As a second step we used crop yield simulations of seven models for the same four crop types of the Intersectoral Impact Model Intercomparison (ISI-MIP) and The Agricultural Model Intercomparison and Improvement Project (AgMIP) fast track (Rosenzweig et al., 2014) obtained from the open-access impact model data archive of ISI-MIP (http://esg.pik-potsdam.de/). These models were driven by output data from five climate models here for the RCP 8.5 pathway, including the suite of processes related to "CO_2- fertilization" for the future period 2070-2099 (modified carboxylation, and in some models reduced stomatal closure). Note that the seven models: EPIC-BOKU (in ISI-MIP/AgMIP fast track refer to the name "EPIC"), GEPIC, GAEZ-IMAGE, LPJ-GUESS, LPJmL, pDSSAT, PEGASUS which took part in the ISI-MIP/AgMIP fast track, also participated in this GGCMI phase 1 study (model details are listed in *SI Appendix* Tables A.1-5), except the GAEZ-IMAGE model.

2.2 Crop masks

Four crop masks were used to aggregate simulated gridded yields: MIRCA2000 (Portmann et al., 2010), Iizumi (Iizumi et al., 2014), Ray (Ray et al., 2012), and SPAM2005 (You et al., 2014). Data sources and main characteristics of the original cropping system data sets were summarized in Table 2.

Table 2

Major features of the four harvested area data sets applied for aggregation

Feature	MIRCA2000	Iizumi	SPAM2005	Ray
Harvested area based on	Monfreda et al. (2008) - with modifications, circa 2000	Monfreda et al. (2008) - circa 2000	FAOSTAT, AGROMAPS and own sub-national data collection, circa 2005	Sub-national data collection (70% to 90%) 1961-2008
National areas	ESRI 2004	Dominant country code per 0.5° grid cell	Same national total areas as in MIRCA2000 (You et al., 2014)	As in Ramankutty et al. (2008)

N° of crops covered	26 crop classes	Maize, soybean, wheat, and rice	20 major crops	Maize, soybean, wheat, and rice
Original resolution	5 arc minute, 0.083° (~10km)	67.5 arc minute, 1.125° (~120km)	5 arc minute, 0.083° (~10km)	5 arc minute, 0.083° (~10km)
Irrigation data based on	Global Map of Irrigation Areas v.4 (Siebert et al., 2007, 2005), AQUASTAT national data	None	Global Map of Irrigation Areas v.5 (Siebert et al., 2007, 2005)	None
Cropland extent based on	Ramankutty et al. (2008)	Ramankutty et al. (2008)	Ramankutty et al. (2008)	Ramankutty et al. (2008)
Data inclusion method	Collection of statistical data and literature	Yield estimation model	Cross entropy approach with spatial allocation model optimization	Administrative bottom-up statistical data inclusion

All four data products were based on the cropland extent (ha) per grid cell by Ramankutty et al. (2008), who merged sub-national and national inventory data with two global satellite based land cover products. MIRCA2000 and Iizumi rely on the harvested area data of Monfreda et al. (2008) who used about 50% of sub-national and also FAO-based national data averaged over the time period 1997-2003. SPAM2005 is the update of the former SPAM2000 data set, wherein the share of sub-national data collection for harvested area was about 50% and Ray's share of that was 70-90% - the rest of both had been complemented with FAO national data as well. MIRCA2000, Iizumi, and SPAM2005 report static harvested area data per grid cell (circa 2000 or 2005) whereas Ray provides a dynamic annual time-series (1961-2008). MIRCA2000 and SPAM2005 independently report the spatial distribution of irrigated and rainfed harvested areas (ha) per crop type, which is an important feature for crop modeling and aggregation but are based on different baseline years (2000 vs. 2005). The Iizumi and Ray data sets do not further distinguish harvested areas into irrigated and rainfed fractions. The four data sets display differences in spatial patterns of harvested area as highlighted by Fig. 1 for maize (for the other crops see *SI Appendix*, Fig. B.1-4)

{Placeholder Figure 1}

2.3 Pre-processing the crop masks

The Iizumi data set, originally reported at a spatial resolution of 1.125°, was interpolated to 0.5°. MIRCA2000, SPAM2005, and Ray originally provided data at 5 arc minutes resolutions, which we aggregated to 0.5°. The original information on cropland extent and harvested area around the year

2000 from MIRCA2000, Iizumi, and SPAM2005 data sets, were kept constant and used to aggregate the simulated yields for the time period 1980-2010. The original Ray data set covered all simulated years up to 2008 and the aggregated yield time series used for this analysis thus spanned only the years 1980-2008. All aggregations with SPAM2005 and MIRCA2000 were performed with their own shares of rainfed and irrigated areas. In the case of the Ray and Iizumi data sets, their harvested area per grid cell were split into irrigated and rainfed fractions using MIRCA2000's relative shares for a given crop in each 0.5° grid cell. Grid cells, for which MIRCA2000 specifies no harvested area for the crop of interest, were assumed to be without irrigation if they contained crops in the original Ray or Iizumi data sets.

2.4 Aggregating gridded yield data

The GGCMs simulations provided crop yield data in tons of dry matter per hectare (t/ha) for four crop types under fully rainfed and fully irrigated conditions in annual time steps within the time period 1980-2010. These grid cell-specific yield estimates have been aggregated to time series at three spatial scales: global, country, and food production unit (FPU, major river basins crossed with countries) (Cai and Rosegrant, 2002) using the four crop masks as weights in the averaging (equation 1):

$$yield_{aggregated} = \frac{\sum_{i=1}^{n} yield_{i\_i} * area\_irrigated_i + \sum_{i=1}^{n} yield_{i\_r} * area\_rainfed_i}{\sum_{i=1}^{n}(area\_irrigated_i + area\_rainfed_i)}$$

i: any grid cell in the aggregation unit

n: number of grid cells in the aggregation unit

$yield_{i\_i}$: simulated yield (t/ha) under full irrigated conditions in grid cell i

$yield_{i\_r}$: simulated yield (t/ha) under rainfed conditions in grid cell i

$area\_irrigated_i$: irrigated harvested area (ha) in grid cell i

$area\_rainfed_i$: rainfed harvested area (ha) in grid cell i

To derive the productivity (t/ha) per year and aggregation unit, each rainfed yield, simulated by the models in a corresponding grid cell, is multiplied with the rainfed harvested area. The same

177 procedure was carried out for the irrigated yields. Then the sum of all rainfed and irrigated
178 production is divided by the total sum of harvested area reported by the individual data sets of that
179 spatial aggregation unit, resulting in the aggregated mean yield (t/ha) per year and aggregation unit.
180 Grid cells were assigned to countries according to the boundary information of Global Administrative
181 Areas (GADM-0, http://gadm.org/), assigning grid cells to the country that has the largest area share
182 in that grid cell. Here we used information on crop specific harvested areas, which can be larger than
183 the physical cropland extent in multiple cropping systems with several harvests per year, which was
184 accounted for in the harvested area data sets. The GGCMs simulated only a single growing period per
185 grid cell, which we assume to be representative for the different growing periods due to current
186 state of implementation of cropping management systems in the models.
187 For an assessment of aggregation uncertainties in projections of future changes in crop productivity,
188 simulated gridded future yields of the ISI-MIP/AgMIP fast track are aggregated to country scale by
189 three different time slices (1961, 1984 and 2008) of the Ray data set.
190 In order to quantify the differences between the different crop mask aggregations, we display
191 absolute (t/ha) and relative (%) differences between yield aggregated with each of the four masks:
192 MIRCA2000 (further abbreviated as MIRCA), Ray, Iizumi, and SPAM2005 (in the following abbreviated
193 as SPAM) for selected regions/countries as well as by computing the yield time series differences
194 over time. The correlation coefficients between the differently aggregated time series were used to
195 describe how yield aggregates of individual years are affected by the different crop masks and how
196 this affects variability over time. If all years were affected equally, aggregated yield time series differ
197 in their mean but are highly correlated. Data analysis was conducted in R (R Development Core
198 Team, 2014), using the standard Pearson correlation (Becker et al., 1988).

199 **3. Results**

200 The different crop masks lead to different yield estimates for individual years at all spatial scales
201 (global, national, and FPU). The mean relative differences among aggregated global yields reach up to
202 6 % for maize, 17 % for wheat, 14 % for rice, and 10% for soybean across the different crop models

(further details at bottom of the Tables 3-6). The ranges depended on the heterogeneity of the simulated spatial yield patterns by the GGCMs and how strongly opposing deviations in different regions compensate each other. The aggregation with different crop masks also affects the simulated temporal dynamics, with minimum correlation coefficients between the global aggregated yield time series of r=0.77 for maize, r=0.85 for wheat, r=0.64 for rice, and r=0.28 for soybean (Tables 3-6).

Across 208 countries, 14 GGCMs, and 31 years, aggregation induced differences between nationally aggregated yield estimates for the four crop types can be very large (>10 DM t /ha), but the majority is below 10% of relative difference (<0.3 DM t/ha in absolute terms). The aggregations with Ray show least differences to aggregations with MIRCA, whereas SPAM-based aggregations show strongest differences to MIRCA, Iizumi, and Ray-based aggregations (Fig. 2). Largest relative differences in yield sets can be found for soybean especially in comparison of SPAM to each of the other three aggregated sets. Aggregated maize yield are least affected by the aggregation uncertainty.

{Placeholder figure 2}

When accounting for differences in total crop area, e.g. when looking at differences in production (t) rather than in productivity (t/ha), the relative differences between country scale aggregations are even stronger (Fig. C in the *SI Appendix*). This is caused by differences in quantity and spatial pattern of the harvested area data set applied for the aggregations. At the national level, the crop cover mask can be of greater importance. In the Tables 3-6, the effects of different aggregations on country scale are displayed for the top-ten producer (for all countries and the four crops Tables D.1-4 in the *SI Appendix*). Differences over the 31 years are shown as the percentage minimum and maximum mean relative difference between the aggregations with Ray, Iizumi, SPAM, and MIRCA-based aggregation. Differences in temporal dynamics induced by the different crop masks applied for the aggregation are shown by the minimum correlation coefficient (r) between aggregated national time series (one per GGCM). Countries were ranked by their share on global production as averaged over the years 2009-2013 (FAO, 2014).

Table 3: Lowest and highest values of mean relative difference (%) and the lowest correlation coefficient (r) between the aggregated maize yield time series (t/ha) calculated from the 14 models, during the AgMERRA time period, aggregated for the top-10 producer countries with one harvested area data set in relation to the aggregation with each of the other three masks (see more detailed results for all countries in *SI Appendix* Table D.1).

maize top-10 producer countries	lowest value of relative difference (%)	masks lowest value of relative difference	highest value of relative difference (%)	masks highest value of relative difference	minimum correlation (r)	masks minimum correlation	Share on global production (%)
USA	-3	SPAM-MIRCA	2	Ray-MIRCA	0.98	Ray-Iizumi	35.74
China	-11	SPAM-MIRCA	8	Ray-SPAM	0.94	Ray-SPAM	21.54
Brazil	-9	SPAM-MIRCA	7	Ray-SPAM	0.95	Ray-Iizumi	7.04
Argentina	-7	Iizumi-MIRCA	10	Ray-Iizumi	0.93	Ray-Iizumi	2.54
Mexico	-14	SPAM-MIRCA	17	Ray-SPAM	0.71	Ray-SPAM	2.38
India	-21	SPAM-MIRCA	38	Ray-SPAM	0.56	MIRCA-Ray	2.38
Ukraine	-11	Iizumi-MIRCA	20	Ray-SPAM	0.96	Iizumi-SPAM	2.18
Indonesia	-8	Iizumi-MIRCA	6	Ray-MIRCA	0.85	Iizumi-SPAM	2.06
France	-20	Iizumi-SPAM	28	SPAM-MIRCA	0.95	MIRCA-Iizumi	1.70
South Africa	-37	SPAM-MIRCA	67	Iizumi-SPAM	0.75	MIRCA-SPAM	1.34
global	-5	Ray-Iizumi	5	Iizumi-MIRCA	0.77	MIRCA-Ray	100

Of the top-10 maize producers (United States, China, Brazil, Argentina, Mexico, India, Ukraine, Indonesia, France, and South Africa) - South Africa, India, and France show stronger sensitivity to the choice of the aggregation mask, while the USA (*SI Appendix* Fig. F.3) is less sensitive to the choice of crop mask (for all countries see *SI Appendix* Table D.1). Of the top-10 maize producers, yield simulations can be strongly affected by the national aggregation mask by up to 67% (South Africa), 38% (India) or 28% (France, Fig. 3). Individual years can be affected more strongly, so that the correlation between the MIRCA-based aggregated time series and the ones obtained with the Ray mask can be low, as in India (r=0.56), while the correlation is not necessarily low if there are stronger differences in mean yields (e.g. France with minimum r=0.95).

{Placeholder figure 3}

From the top-10 wheat producer countries (Table 4) Canada with -28-41% has the largest span of relative yield difference as well as a low correlation coefficient of r=0.41 (Iizumi-SPAM). For Pakistan, differences in mean yield of up to 43% can be observed for the MIRCA-based aggregation compared to the one with Iizumi. Only the mean relative difference between aggregated yield sets for Russia,

United States, France, and Germany are about 15% or less. For the case of wheat productivity in Russia low differences in yields are shown but the correlation coefficient reaches as low values as r=0.05 displaying the larger deviations of temporal patterns in aggregated yield sets (MIRCA-SPAM).

Table 4: Lowest and highest values of mean relative difference (%) and the lowest correlation coefficient (r) between the aggregated wheat yield time series (t/ha) calculated from the 14 models, during the AgMERRA time period, aggregated for the top-10 producer countries with one harvested area data set in relation to the aggregation with each of the other three masks (see more detailed results for all countries in (*SI Appendix* Table D.2).

wheat top-10 producer countries	lowest value of relative difference (%)	masks lowest value of relative difference	highest value of relative difference (%)	masks highest value of relative difference	minimum correlation (r)	masks minimum correlation	Share on global production (%)
China	-19	SPAM-MIRCA	19	Iizumi-SPAM	0.82	SPAM-MIRCA	17.26
India	-16	SPAM-MIRCA	33	Iizumi-SPAM	0.89	Iizumi-SPAM	12.77
USA	-8	Iizumi-MIRCA	7	Ray-Iizumi	0.77	Iizumi-SPAM	8.61
Russia	-6	Iizumi-SPAM	6	Iizumi-SPAM	0.05	SPAM-MIRCA	7.29
France	-5	Iizumi-SPAM	6	Ray-Iizumi	0.85	Iizumi-MIRCA	5.60
Canada	-28	Ray-SPAM	41	SPAM-MIRCA	0.41	Iizumi-SPAM	4.09
Australia	-21	Iizumi-SPAM	16	SPAM-MIRCA	0.87	Iizumi-SPAM	3.62
Pakistan	-19	SPAM-MIRCA	43	Iizumi-MIRCA	0.79	SPAM-MIRCA	3.52
Germany	-4	Iizumi-MIRCA	5	Ray-Iizumi	0.94	MIRCA-Ray	3.50
Turkey	-17	Iizumi-SPAM	15	SPAM-MIRCA	0.72	MIRCA-Ray	3.05
global	-17	SPAM-MIRCA	10	Ray-SPAM	0.85	MIRCA-Ray	100

In the case of rice productivity (Table 5), relative differences between aggregations sets for Indonesia and Brazil are below 10%. Indonesia has fairly high correlation across all masks pairings but for Brazil the correlation between the MIRCA and Ray-based aggregations is as low as r=0.32. Rice yields for Vietnam, Philippines, Thailand, and Japan show very strong relative differences between aggregated yield sets. For rice in Vietnam also the temporal dynamics are affected by the choice of aggregation mask, reflected by a very low correlation coefficient of r=0.13 when comparing MIRCA- to SPAM-based aggregations.

Table 5: Lowest and highest values of mean relative difference (%) and the lowest correlation coefficient (r) between the aggregated rice yield time series (t/ha) calculated from 11 models, during the AgMERRA time period, aggregated for the top-10 producer countries with one harvested area

data set in relation to the aggregation with each of the other three masks (see more detailed results for all countries in *SI Appendix* Table D.3). Note that the models PEGASUS, PAPSIM, and EPIC-TAMU did not simulate rice.

rice top-10 producer countries	lowest value of relative difference (%)	masks lowest value of relative difference	highest value of relative difference (%)	masks highest value of relative difference	minimum correlation (r)	masks minimum correlation	Share on global production (%)
China	-25	Iizumi-MIRCA	14	SPAM-MIRCA	0.71	MIRCA-Ray	27.99
India	-10	Iizumi-SPAM	13	SPAM-MIRCA	0.88	MIRCA-Ray	20.97
Indonesia	-5	Iizumi-MIRCA	4	Ray-SPAM	0.95	Iizumi-SPAM	9.36
Bangladesh	-15	Iizumi-SPAM	17	SPAM-MIRCA	0.97	MIRCA-SPAM	6.97
Vietnam	-33	Iizumi-SPAM	42	SPAM-MIRCA	0.13	MIRCA-SPAM	5.81
Thailand	-29	Iizumi-SPAM	35	SPAM-MIRCA	0.78	Ray-SPAM	4.97
Myanmar	-11	Iizumi-SPAM	10	Ray-SPAM	0.92	MIRCA-SPAM	4.18
Philippines	-33	Iizumi-SPAM	38	SPAM-MIRCA	0.77	Ray-SPAM	2.37
Brazil	-9	Ray-Iizumi	8	Iizumi-SPAM	0.32	MIRCA-Ray	1.69
Japan	-18	Ray-Iizumi	51	Iizumi-MIRCA	0.79	MIRCA-Ray	1.48
global	-14	Iizumi-SPAM	11	SPAM-MIRCA	0.64	MIRCA-Ray	100

For soybean several countries show large relative differences attributed to the crop mask and the modelled yield patterns across the country. For soybean in Bolivia the relative difference between the Ray and the SPAM-based aggregation reach 427%, for Paraguay 82% between Iizumi- and SPAM-based aggregations, followed by India with 48% relative yield difference between the Ray- and the SPAM-based aggregation. China and the United States show the lower sensitivity to the crop mask applied with ranging around 10% relative difference between the different aggregated yield sets. Although soybean yields of Brazil show relatively low sensitivity to the aggregation mask effects with 23% as maximum relative difference, but the correlation coefficient of r=0.07 between the Ray- to SPAM-based aggregation is very low, displaying little agreement in temporal pattern between the time series. Temporal dynamics of soybean productivity in Uruguay, Canada, and India are greatly affected by the aggregation mask and can reach even negative correlation coefficients.

Table 6: Lowest and highest values of mean relative difference (%) and the lowest correlation coefficient (r) between the aggregated soybean yield time series (t/ha) calculated from 13 models, during the AgMERRA time period, aggregated for the top-10 producer countries with one harvested area data set in relation to the aggregation with each of the other three masks (see more detailed

results for all countries in *SI Appendix* Table D.4). Note that the model EPIC-TAMU did not simulate soybean.

soybean top-10 producer countries	lowest value of relative difference (%)	masks lowest value of relative difference	highest value of relative difference (%)	masks highest value of relative difference	minimum correlation (r)	masks minimum correlation	Share on global production (%)
USA	-4	Ray-SPAM	9	Ray-MIRCA	0.91	Ray-SPAM	34.52
Brazil	-8	Iizumi-MIRCA	23	Ray-Iizumi	0.07	Ray-SPAM	27.48
Argentina	-22	Ray-Iizumi	25	Iizumi-MIRCA	0.8	Ray-Iizumi	17.51
China	-8	SPAM-MIRCA	14	Iizumi-SPAM	0.83	Ray-SPAM	5.53
India	-13	SPAM-MIRCA	48	Ray-SPAM	-0.08	Ray-MIRCA	4.85
Paraguay	-41	SPAM-MIRCA	82	Iizumi-SPAM	0.83	SPAM-MIRCA	2.61
Canada	-16	SPAM-MIRCA	20	Ray-SPAM	-0.23	SPAM-MIRCA	1.77
Uruguay	-16	Ray-SPAM	27	Iizumi-SPAM	-0.01	Ray-SPAM	0.88
Ukraine	-9	SPAM-MIRCA	12	Ray-SPAM	0.82	Ray-SPAM	0.80
Bolivia	-68	SPAM-MIRCA	427	Ray-SPAM	0.45	Ray-SPAM	0.78
global	-6	SPAM-MIRCA	10	Ray-SPAM	0.28	Ray-SPAM	100

The differences due to aggregation can become exceptionally high in countries with pronounced differences in crop-specific harvested area information (*SI Appendix* Tables G.1-2) and where GGCMs simulate heterogeneous yield patterns, as reflecting strong gradients in climatic conditions or crop management practices. Strong yield gradients between grid cells within a country can also derive from model-specific calibration processes of e.g. simulated yields to observations of field experiments or country-specific reference data sets (*SI Appendix* Table A.5). The effect of calibration may even increase the aggregation uncertainty, which is exemplified by maize yield aggregations in Egypt (Fig. 4, *SI Appendix* Fig. E.1). In Egypt almost the entire maize production is irrigated. In Fig.4 we show GGCM simulations of four different models. PEGASUS and PRYSBI2 simulate very heterogeneous yield patterns, whereas pDSSAT assumes more homogeneous and LPJmL simulates very homogeneous yield patterns, assuming national uniform crop production intensities.

{Placeholder figure 4}

In the case of model PRYSBI2, the only area with higher yields is around Port Said, for which only the Iizumi crop mask reports some larger harvested area for maize (Fig. 4, *SI Appendix* G.1-2). PRYSBI2 calibrates several parameters (more details in *SI Appendix* Table A.5) on grid cell level to best match the yields to the Iizumi et al. (2014) yield reference data set in their "default" simulation. Consequently, aggregated PRYSBI2 yields are very low, except when aggregated with the Iizumi crop mask, which results in an aggregated annual yield being up to 250% more productive compared to

306 the other aggregations. For the model PEGASUS, the productive harvested area is located along the
307 Mediterranean coastline. Calibration in PEGASUS consisted in tuning the radiation use efficiency
308 factor (ß) to select a proper crop variety to best match the yield data of Monfreda et al. (2008)
309 according to the Willmott index of agreement. The aggregated national result for PEGASUS's yields
310 shows stronger differences for the SPAM aggregation, which reports less harvested maize areas
311 along the Mediterranean coast line. LPJmL calibrates its parameters: maximum leaf-area-index under
312 unstressed conditions, harvest index, and factor (alpha) for up-scaling leaf-level-photosynthesis to
313 stand level, at country scale, to best match the national yields reported by the FAO. LPJmL thus
314 simulated a very homogeneous yield pattern for irrigated maize in Egypt, as climatic conditions are
315 similarly very hot and dry - but irrigated across the area. The yields of pDSSAT are calibrated to field
316 experiment results. The maize yield pattern of pDSSAT for Egypt is less homogeneous than LPJmL as
317 it takes into account more spatial detail on fertilizer application and other management parameters.
318 Further analysis reveals that sub-regions of larger producing countries, as in individual FPUs of the
319 USA, show a mixed response. Major production areas of the USA along the Mississippi (*SI Appendix*
320 Fig. F.1), the Missouri, and the Ohio River catchments show very little sensitivity to the choice of the
321 crop mask. Other FPUs, such as the Colorado River catchment (*SI Appendix* Fig.F.2) or California,
322 show larger discrepancies between the aggregated yield sets. At the national scale, these regional
323 discrepancies do not show, as the national aggregated productivity is numerically dominated by the
324 major production areas, which show little sensitivity to the choice of the aggregation mask (*SI*
325 *Appendix* Fig. F.3)
326 Assuming static crop masks in the assessments of climate change impacts on agricultural productivity
327 can also strongly affect the projected impact on crop yields. We demonstrate this by aggregating the
328 climate change impact projections on yields of the ISI-MIP/AgMIP fast track (Rosenzweig et al., 2014)
329 with different time slices of the Ray crop mask (years: 1961, 1984, and 2008) as if the assessment had
330 been conducted in these years, assuming 'current' crop masks. We find strong effects on the
331 projected future yield changes in response to climate and elevated atmospheric CO_2 for individual

crops in some countries. Figure 5 shows the differences in projected relative yield changes (percentage change of the period 2070-2099 relative to the 1980-2009 baseline) between the country scale aggregation with the 1961 mask and the aggregation with the two other masks (1984 and 2008) for all seven models that contributed to the ISI-MIP/AgMIP fast track (Rosenzweig et al., 2014). The differences in the five climate projections affect the heterogeneity of simulated yields and thus the sensitivity of aggregated yield changes to the crop mask (bars and whiskers in Fig. 5). For aggregated maize yield projections in India most models show a positive trend with time in projected changes in yields. The projected difference in relative yield change simulated by EPIC-BOKU, GEPIC, and pDSSAT models are considerably higher for the aggregation with Ray's harvested area time slice of 2008 compared to the 1961 as the relative yield change of the aggregated yield with the 1984 mask compared to 1961er. For the case of wheat in Australia the projected yield changes agree quite well, showing only slightly median differences between the time slices used for aggregation. Only the EPIC-BOKU projections show a high variability and maximal difference of yield change of up to -10% with the 2008er in comparison to the 1961 mask but only 4% difference for the 1984 in comparison to the 1961 time slice. This is because the crop-specific harvested area regions in the former case have changed a lot with significant expansion of harvested maize areas in southern India, whereas in Australia the regions have remained roughly similar.

{Placeholder figure 5}

In the case of rice productivity in Brazil, aggregations with the crop mask of 2008 lead to higher difference in yield change projections than the 1984 mask (except for GEPIC) compared to the aggregation with the 1961 time slice. For soybean in Argentina the magnitude of differences in projected yield change are less pronounced between the time-slices' aggregation used but are very different among models as for pDSSAT, and LPJ-GUESS up to 20% but more than 40% for PEGASUS. Differences in climate change impact projections for all other countries of the top-10 producer countries are lower than for those countries displayed in Fig.5.

4. Discussion

We find that differences in crop masks affect not only the mean bias of aggregated yield time series but also the temporal dynamics, resulting in low or even negative correlations between the differently aggregated time series (Tables 3-6, and D.1-8 in the *SI Appendix*). This is of particular concern, as model skill is often determined by comparing temporal dynamics rather than mean yields. Large difference between aggregated yield time series occur, when areas suitable for crop growth (determined by the individual model) are combined with a large harvested area reported by one mask but rather little by another (Fig. 4, *SI Appendix*, Tables G.1-2). Developers of GGCMs need to analyze the spatial variability of their simulations for plausibility. Models that tend to simulate very heterogeneous patterns of crop yields due to calibration, flexible parameter specifications, and assumptions on management practices (e.g. cultivar choice, fertilizer application, sowing dates) were more sensitive to the choice of crop mask (*SI Appendix*, Table A.5). Further differences between the aggregated productivity time series result from the fact, that spatial location of national borders of the various original crop masks are different due to different data products included by the authors (Table 2). When applying publicly available statistics for down-scaling data to a grid cell (as the authors did to produce the harvested area data sets) its accuracy is also limited by the fact, that the historical development of states cannot be well reflected in a timely manner. Also, we assume that each grid cell always belongs to a single country, whereas often the simulated grid cell level results would need to be attributed as fractions to multiple countries. However, since we treat this consistently across the different crop mask data sets used, we consider the resulting error as not relevant in the comparison of the different crop masks in the aggregation process.

The spatial patterns of crop-specific harvested areas as provided by the four data sets here used for aggregation, and the information on where irrigation is applied for these crops is central to large-scale crop modeling. The crop-modelling community requires more complex and updated data on the spatial and temporal dynamics of agricultural production systems. The Ray data set is the only crop mask that is dynamic in time and it also is typically the aggregation mask that shows the largest

differences in the temporal dynamics between the aggregated yield time series (low correlation coefficients). We conclude that each of the four harvested area data sets has its unique features and none can be identified as particularly superior by our study. For particular regions spatial aggregations should be performed with alternative crop masks to assess the effects of aggregation uncertainty and to avoid drawing erroneous conclusions on model skill or projected impacts.

Reporting productivity is what is typically done to communicate or analyze climate change impacts on agriculture (e.g. Müller et al., 2015; Osborne et al., 2013; Wheeler and von Braun, 2013) or to inform land use change models (Müller and Robertson, 2014; Nelson et al. 2014a,b; Schmitz et al., 2014). With some exceptions, as e.g. GLOBIOM (Havlík et al., 2012, 2011) and MAgPIE (Dietrich et al., 2014; Lotze-Campen et al., 2008), these models require information on changes in agricultural productivity aggregated to their simulation units (because of their often coarser resolution, as e.g. national or supra-national regions). General shifts of cropping areas towards higher productive areas are very likely (Beddow and Pardey, 2015) as can be investigated by land use models, which project changes in land use and production as socio-economic responses to changes in agricultural productivity. Future land use uncertainty can also be addressed by aggregating simulated changes in productivity with external land use scenarios as in Pugh et al. (2015) and remain a challenge for further crop modeling studies.

5. Conclusions

This study shows quantitative differences between the aggregated gridded yield time series revealing the uncertainty induced by the aggregation applying differing harvested area data sets. The effects of aggregation uncertainty are the shift of the multi-annual mean national yield and an influence on the variability over time, depending on the heterogeneity of simulated yield patterns by the models and the differences between crop masks. This uncertainty is already significant in global aggregations of grid cell scale yield simulations and can be very large for some aggregation-unit-crop-model-year combinations. Aggregation uncertainty of gridded yields becomes even more important when taking

into account production instead of productivity. For projections of future agricultural production, this aggregation uncertainty will likely be small compared to given uncertainties in future climate change, adaptation options, and capacities. The potentially large differences between different aggregations for individual countries or regions will have to be considered in future model evaluations and also in future crop yield projections. This requires considerable investment for building a transparent method for aggregation. The study also illustrates the need to transition from assuming static harvested areas towards dynamic projections that account for spatial shifts in crop distribution and production induced by changes in social and environmental conditions.

Acknowledgement

We acknowledge the support and data provision by the Agricultural Intercomparison and Improvement Project (AgMIP), the Intersectoral Impact Model Intercomparison Project (ISIMIP), and the contributing modelers. V.P. and C.M. acknowledge financial support from the MACMIT project (01LN1317A) funded through the German Federal Ministry of Education and Research (BMBF). A.A. and T.A.M.P. were supported by the European Commission's 7th Framework Programme under Grant Agreement number 603542 (LUC4C) and by the Helmholtz Association through its research program ATMO. This represents paper number 21 of the Birmingham Institute of Forest Research.

Author's contribution

The research question to this paper has been developed and proposed by the GGCMI coordinators J.E. and C.M. J.E. and J.C. performed the post-processing as aggregating the submitted data from grid cell-level to coarser spatial units. C.M. and V.P. conducted the analysis. V.P. wrote the manuscript with substantial contributions from C.M., P.C., D.R., T.I., J.E, D.D., R.C.I, and C.J. All co-authors provided data to the GGCMI project, discussed, and commented on the manuscript.

References

Anderson, W., You, L., Wood, S., Wood-Sichra, U., Wu, W., 2015. An analysis of methodological and spatial differences in global cropping systems models and maps. Global Ecol Biogeogr 24(2), 180-191, DOI: 10.1111/geb.12243.

Asseng, S., Ewert, F., Martre, P., Rotter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V.V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.K., Muller, C., Naresh Kumar, S., Nendel, C., O/'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stockle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., Zhu, Y., 2015. Rising temperatures reduce global wheat production. Nature Clim. Change 5(2), 143-147, DOI: 10.1038/nclimate2470.

Balkovič, J., van der Velde, M., Schmid, E., Skalský, R., Khabarov, N., Obersteiner, M., Stürmer, B., Xiong, W., 2013. Pan-European crop modelling with EPIC: Implementation, up-scaling and regional crop yield validation. Agric. Syst 120, 61-75, DOI: 10.1016/j.agsy.2013.05.008.

Batjes, N.H., 2006. ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (version 1.1). Report 2006/02 (http://www.isric.org). SRIC- World Soil Information, Wageningen, Netherlands.

Becker, R.A., Chambers, J.M., Wilks, A.R., 1988. The new S language: A programming environment for data analysis and graphics, Wadsworth and Brooks/Cole Advanced Books & Software.USA.

Bondeau, A., Smith, P.C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Müller, C., Reichstein, M., Smith, B., 2007. Modelling the role of agriculture for the 20th century global terrestrial carbon balance. Glob Change Biol 13 (3), 679-706, DOI: 10.1111/j.1365-2486.2006.01305.x.

Beddow, J.M.,Pardey, P.G., 2015. Moving matters: The effect of location on crop production. The J. Econ. Hist. 75(01), 219-249, DOI: doi:10.1017/S002205071500008X.

Cai, X.,Rosegrant, M.W., 2002. Global water demand and supply projections. Water Int 27 (2), 159-169, DOI: 10.1080/02508060208686989.

Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R., Chhetri, N., 2014. A meta-analysis of crop yield under climate change and adaptation. Nature Clim. Change 4(4), 287-291, DOI: 10.1038/nclimate2153.

Cosby, B.J., Hornberger, G.M., Clapp, R.B., Ginn, T.R., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water Resour. Res 20 (6), 682-690, DOI: 10.1029/WR020i006p00682.

de Wit, A.J.W., van Diepen, C.A., 2008. Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs. Int. J. Appl. Earth Obs. Geoinf. 10(4), 414-425, DOI: 10.1016/j.jag.2007.10.004.

Deryng, D., Conway, D., Ramankutty, N., Price, J., Warren, R., 2014. Global crop yield response to extreme heat stress under multiple climate change futures. Environ. Res. Lett. 9 (3), 034011, DOI: 10.1088/1748-9326/9/3/034011

Deryng, D., Sacks, W.J., Barford, C.C., Ramankutty, N., 2011. Simulating the effects of climate and agricultural management practices on global crop yield. Global Biogeochem. Cy, 25 (2), GB2006, DOI: 10.1029/2009GB003765.

Dietrich, J.P., Schmitz, C., Lotze-Campen, H., Popp, A.,Müller, C., 2014. Forecasting technological change in agriculture—an endogenous implementation in a global land use model. Technol. Forecast Soc. 81, 236-249, DOI: doi.org/10.1016/j.techfore.2013.02.003.

Dobos, E., 2006. Albedo. In Encyclopedia of Soil Science, Second Edition., 64-66.

Drewniak, B., Song, J., Prell, J., Kotamarthi, V.R., Jacob, R., 2013. Modeling agriculture in the community land model. Geosci. Model Dev.6 (2), 495-515, DOI: 10.5194/gmd-6-495-2013.

Eitelberg, D.A., van Vliet, J., Verburg, P.H., 2015. A review of global potentially available cropland estimates and their consequences for model-based assessments. Glob. Change Biol. 21 (3), 1236-1248, DOI: 10.1111/gcb.12733.
Elliott, J., Kelly, D., Chryssanthacopoulos, J., Glotter, M., Jhunjhnuwala, K., Best, N., Wilde, M., Foster, I., 2014. The parallel system for integrating impact models and sectors (PSIMS). Environ. Model. Softw. 62, 509-516, DOI: 10.1016/j.envsoft.2014.04.008.
Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K.J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R.C., Mueller, N.D., Ray, D.K., Rosenzweig, C., Ruane, A.C., Sheffield, J., 2015. The global gridded crop model intercomparison: Data and modeling protocols for phase 1 (v1.0). Geosci. Model Dev.8(2), 261-277, DOI: 10.5194/gmd-8-261-2015.
FAO, 2014. FAOSTAT: Agricultural production. from: http://faostat.fao.org/, date: 2014/09/29.
Farquhar, G.D., von Caemmerer, S., Berry, J.A., 1980. A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. Planta 149 (1), 78-90, DOI: 10.1007/BF00386231.
Fila, G., Bellocchi, G., Acutis, M., Donatelli, M., 2003. Irene: A software to evaluate model performance. Eur. J. Agron. 18(3), 369-372, DOI: 10.1016/S1161-0301(02)00129-6.
Fischer, G., Nachtergaele, F., Prieler, S., Teixeira, E., Tóth, G., van Velthuizen, H., Verelst, L., Wiberg, D., 2012. Global agro-ecological zones (GAEZ v3. 0) - Model documentation. International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. Food and Agriculture Organization of the United Nations (FAO) Rome, Italy.
Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H.T., Verelst, L., Wiberg, D., 2008. Global agro-ecological zones assessment for agriculture (GAEZ 2008).
Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C., Hansen, M., Gong, P., Abdel Aziz, S., Cipriani, A., Cumani, R., Cecchi, G., Conchedda, G., Ferreira, S., Gomez, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A., Ortner, S., Rajak, D.R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhov, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W., van der Velde, M., Dunwoody, A., Kraxner, F., Obersteiner, M., 2015. Mapping global cropland and field size. Glob. Change Biol., DOI: 10.1111/gcb.12838.
Hall, F.G., Brown de Colstoun, E., Collatz, G.J., Landis, D., Dirmeyer, P., Betts, A., Huffman, G.J., Bounoua, L., Meeson, B., 2006. Islscp initiative ii global data sets: Surface boundary conditions and atmospheric forcings for land-atmosphere studies. J. Geophys. Res. : Atmospheres 111(D22), D22S01, DOI: 10.1029/2006JD007366.
Havlík, P., Schneider, U.A., Schmid, E., Böttcher, H., Fritz, S., Skalský, R., Aoki, K., Cara, S.D., Kindermann, G., Kraxner, F., Leduc, S., McCallum, I., Mosnier, A., Sauer, T.,Obersteiner, M., 2011. Global land-use implications of first and second generation biofuel targets. Energ. Policy 39(10), 5690-5702, DOI: 10.1016/j.enpol.2010.03.030.
Havlík, P., Valin, H., Mosnier, A., Obersteiner, M., Baker, J.S., Herrero, M., Rufino, M.C.,Schmid, E., 2012. Crop productivity and the global livestock sector: Implications for land use change and greenhouse gas emissions. American Journal of Agricultural Economics, DOI: 10.1093/ajae/aas085.
Lotze-Campen, H., Müller, C., Bondeau, A., Rost, S., Popp, A.,Lucht, W., 2008. Global food demand, productivity growth, and the scarcity of land and water resources: A spatially explicit mathematical programming approach. Agr. Econ. 39(3), 325-338, DOI: 10.1111/j.1574-0862.2008.00336.x.
Iizumi, T., Yokozawa, M., Sakurai, G., Travasso, M.I., Romanenkov, V., Oettli, P., Newby, T., Ishigooka, Y., Furuya, J., 2014. Historical changes in global yields: Major cereal and legume crops from 1982 to 2006. Global Ecol. Biogeogr. 23 (3), 346-357, DOI: 10.1111/geb.12120.
Izaurralde, R.C., McGill, W.B., Williams, J.R., 2012. Chapter 17 - development and application of the epic model for carbon cycle, greenhouse gas mitigation, and biofuel studies. Managing agricultural greenhouse gases. Liebig, Mark A. , Franzluebbers, Alan J. and Follett, Ronald F. San Diego, Academic Press: 293-308.

Izaurralde, R.C., Williams, J.R., McGill, W.B., Rosenberg, N.J., Jakas, M.C.Q., 2006. Simulating soil c dynamics with epic: Model description and testing against long-term data. Ecol. Model. 192 (3–4), 362-384, DOI: 10.1016/j.ecolmodel.2005.07.010.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. Eur. J. Agron. 18(3-4), 235-265, DOI: 10.1016/s1161-0301(02)00107-7.

Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N.G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J.P., Silburn, M., Wang, E., Brown, S., Bristow, K.L., Asseng, S., Chapman, S., McCown, R.L., Freebairn, D.M., Smith, C.J., 2003. An overview of APSIM, a model designed for farming systems simulation. Eur. J. Agron. 18 (3–4), 267-288, DOI: 10.1016/S1161-0301(02)00108-9.

Lawrence, D., Slater, A., 2008. Incorporating organic soil into a global climate model. Clim. Dyn. 30(2-3), 145-160, DOI: 10.1007/s00382-007-0278-1.

Leff, B., Ramankutty, N., Foley, J.A., 2004. Geographic distribution of major crops across the world. Global Biogeochem. Cy. 18 (1), DOI: 10.1029/2003GB002108.

Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., Smith, B., 2013. Implications of accounting for land use in simulations of ecosystem carbon cycling in africa. Earth Syst. Dynam. 4 (2), 385-407, DOI: 10.5194/esd-4-385-2013.

Liu, J., Williams, J.R., Zehnder, A.J.B., Yang, H., 2007. Gepic - modelling wheat yield and crop water productivity with high resolution on a global scale. Agric. Syst. 94 (2), 478-493, DOI: 10.1016/j.agsy.2006.11.019.

Liu, W., Yang, H., Folberth, C., Wang, X., Luo, Q., Schulin, R., 2016. Global investigation of impacts of pet methods on simulating crop-water relations for maize. Agr. Forest Meteorol. 221, 164-175, DOI: 10.1016/j.agrformet.2016.02.017.

Monfreda, C., Ramankutty, N., Foley, J.A., 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Global Biogeochem. Cy. 22 (1), GB1022, DOI: 10.1029/2007GB002947.

Mueller, N.D., Gerber, J.S., Johnston, M., Ray, D.K., Ramankutty, N.,Foley, J.A., 2012. Closing yield gaps through nutrient and water management. Nature 490 (7419), 254-257, DOI: 10.1038/nature11420.

Müller, C., Elliott, J., Chryssanthacopoulos, J., Deryng, D., Folberth, C., Pugh, T.A.M.,Schmid, E., 2015. Implications of climate mitigation for future agricultural production. Environ. Res. Lett. 10(12), 125004, DOI: 10.1088/1748-9326/10/12/125004.

Müller, C., Robertson, R.D., 2014. Projecting future crop productivity for global economic modeling. Agr. Econ. 45 (1), 37-50, DOI: 10.1111/agec.12088.

Nelson, G.C., Valin, H., Sands, R.D., Havlík, P., Ahammad, H., Deryng, D., Elliott, J., Fujimori, S., Hasegawa, T., Heyhoe, E., Kyle, P., Von Lampe, M., Lotze-Campen, H., Mason d'Croz, D., van Meijl, H., van der Mensbrugghe, D., Müller, C., Popp, A., Robertson, R., Robinson, S., Schmid, E., Schmitz, C., Tabeau, A.,Willenbockel, D., 2014a. Climate change effects on agriculture: Economic responses to biophysical shocks. Proc. Natl. Acad. Sci. USA 111(9), 3274-3279, DOI: 10.1073/pnas.1222465110.

Nelson, G.C., van der Mensbrugghe, D., Ahammad, H., Blanc, E., Calvin, K., Hasegawa, T., Havlik, P., Heyhoe, E., Kyle, P., Lotze-Campen, H., von Lampe, M., Mason d'Croz, D., van Meijl, H., Müller, C., Reilly, J., Robertson, R., Sands, R.D., Schmitz, C., Tabeau, A., Takahashi, K., Valin, H., Willenbockel, D., 2014b. Agriculture and climate change in global scenarios: Why don't the models agree. Agr. Econ. 45 (1), 85-101, DOI: 10.1111/agec.12091.

Osborne, T., Rose, G.,Wheeler, T., 2013. Variation in the global-scale impacts of climate change on crop productivity due to climate model uncertainty and adaptation. Agr. Forest Meteorol. 170, 183-194, DOI: 10.1016/j.agrformet.2012.07.006.

Portmann, F.T., Siebert, S., Döll, P., 2010. Mirca2000—global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. Global Biogeochem. Cy. 24 (1), GB1011, DOI: 10.1029/2008GB003435.

Pugh, T.A.M., Arneth, A., Olin, S., Ahlström, A., Bayer, A.D., Goldewijk, K.K., Lindeskog, M.,Schurgers, G., 2015. Simulated carbon emissions from land-use change are substantially enhanced by accounting for agricultural management. Environ. Res. Lett. 10(12), DOI: 10.1088/1748-9326/10/12/124008

R Development Core Team (2014). R: A language and environment for statistical computing. Computing, R Foundation for Statistical. Vienna, Austria.

Ramankutty, N., Evan, A.T., Monfreda, C., Foley, J.A., 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. Global Biogeochem. Cy. 22 (1), GB1003, DOI: 10.1029/2007GB002952.

Ray, D.K., Ramankutty, N., Mueller, N.D., West, P.C., Foley, J.A., 2012. Recent patterns of crop yield growth and stagnation. Nat. Commun. 3, DOI: 10.1038/ncomms2296.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T.A.M., Schmid, E., Stehfest, E., Yang, H., Jones, J.W., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proc. Natl. Acad. Sci. U.S.A. 111 (9), 3268-3273, DOI: 10.1073/pnas.1222463110.

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., 2013. The agricultural model intercomparison and improvement project (AgMIP): Protocols and pilot studies. Agr. and Forest Meteorol. 170, 166-182, DOI: 10.1016/j.agrformet.2012.09.011.

Rötter, R.P., Palosuo, T., Kersebaum, K.C., Angulo, C., Bindi, M., Ewert, F., Ferrise, R., Hlavinka, P., Moriondo, M., Nendel, C., Olesen, J.E., Patil, R.H., Ruget, F., Taká, J., Trnka, M., 2012. Simulation of spring barley yield in different climatic zones of northern and central europe: A comparison of nine crop models. Field Crop. Res. 133, 23-36, DOI: 10.1016/j.fcr.2012.03.016.

Roux, S., Brun, F., Wallach, D., 2014. Combining input uncertainty and residual error in crop model predictions: A case study on vineyards. Eur. J. Agron. 52 (Part B), 191-197, DOI:10.1016/j.eja.2013.09.008.

Ruane, A.C., Goldberg, R., Chryssanthacopoulos, J., 2015. Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. Agr. and Forest Meteorol. 200, 233-248, DOI: 10.1016/j.agrformet.2014.09.016.

Sacks, W.J., Deryng, D., Foley, J.A.,Ramankutty, N., 2010. Crop planting dates: An analysis of global patterns. Global Ecol. Biogeogr. 19 (5), 607-620, DOI: 10.1111/j.1466-8238.2010.00551.x.

Sakurai, G., Iizumi, T., Nishimori, M., Yokozawa, M., 2014. How much has the increase in atmospheric CO_2 directly affected past soybean production? Scientific Reports 4, 4978, DOI: 10.1038/srep04978.

Schaap, M.G.,Bouten, W., 1996. Modeling water retention curves of sandy soils using neural networks. Water Resour. Res. 32 (10), 3033-3040, DOI: 10.1029/96WR02278.

Schmitz, C., van Meijl, H., Kyle, P., Nelson, G.C., Fujimori, S., Gurgel, A., Havlik, P., Heyhoe, E., d'Croz, D.M., Popp, A., Sands, R., Tabeau, A., van der Mensbrugghe, D., von Lampe, M., Wise, M., Blanc, E., Hasegawa, T., Kavallari, A., Valin, H., 2014. Land-use change trajectories up to 2050: Insights from a global agro-economic model comparison. Agr. Econ. 45 (1), 69-84, DOI: 10.1111/agec.12090.

See, L., Fritz, S., You, L., Ramankutty, N., Herrero, M., Justice, C., Becker-Reshef, I., Thornton, P., Erb, K., Gong, P., Tang, H., van der Velde, M., Ericksen, P., McCallum, I., Kraxner, F., Obersteiner, M., 2015. Improved global cropland data as an essential ingredient for food security. Glob. Food Sec. 4, 37-45, DOI: 10.1016/j.gfs.2014.10.004.

Siebert, S., Döll, P., Feick, S., Frenken, K., Hoogeveen, J., 2007. Global map of irrigation areas version 4.0.1. University of Frankfurt (Main), Germany, and FAO, Rome, Italy.

Siebert, S., Döll, P., Hoogeveen, J., Faures, J.M., Frenken, K., Feick, S., 2005. Development and validation of the global map of irrigation areas. Hydrol. Earth Syst. Sci. 9 (5), 535-547, DOI: 10.5194/hess-9-535-2005.

Smith, B., Prentice, I.C., Sykes, M.T., 2001. Representation of vegetation dynamics in the modelling of terrestrial ecosystems: Comparing two contrasting approaches within european climate space. Global Ecol. Biogeogr. 10 (6), 621-637, DOI: 10.1046/j.1466-822X.2001.t01-1-00256.x.

633 Thoning, K.W., Tans, P.P., Komhyr, W.D., 1989. Atmospheric carbon dioxide at Mauna Loa
634 observatory: 2. Analysis of the NOAA GMCC data, 1974–1985. J. Geophys. Res. : Atmospheres 94
635 (D6), 8549-8565, DOI: 10.1029/JD094iD06p08549.
636 USDA,NRCS, 2015. Web soil survey. from: http://websoilsurvey.nrcs.usda.gov/ (accessed:
637 09.11.2015).
638 Van Genuchten, M.T., Leij, F., Lund, L. 1992. Indirect methods for estimating the hydraulic properties
639 of unsaturated soils, University of California, Riverside, USA.
640 van Wart, J., van Bussel, L.G.J., Wolf, J., Licker, R., Grassini, P., Nelson, A., Boogaard, H., Gerber, J.,
641 Mueller, N.D., Claessens, L., van Ittersum, M.K., Cassman, K.G., 2013. Use of agro-climatic zones to
642 upscale simulated crop yield potential. Field Crop. Res. 143, 44-55, DOI: 10.1016/j.fcr.2012.11.023.
643 Waha, K., van Bussel, L.G.J., Müller, C., Bondeau, A., 2012. Climate-driven simulation of global crop
644 sowing dates. Global Ecol. Biogeogr. 21 (2), 247-259, DOI: 10.1111/j.1466-8238.2011.00678.x.
645 Wheeler, T.,von Braun, J., 2013. Climate change impacts on global food security. Science 341(6145),
646 508-513, DOI: 10.1126/science.1239402.
647 Williams, J.R., 1995. The EPIC model. Computer models of watershed hydrology. Littleton, CO**:** 909-
648 1000.
649 Willmott, C., Ackleson, S., Davis, R., Feddema, J., Klink, K., Legates, D., O'Donnell, J., Rowe, C., 1985.
650 Statistics for the evaluation and comparison of models. J. Geophys. Res. 90 (C5), 8995-9005, DOI:
651 10.1029/JC090iC05p08995.
652 Wu, X., Vuichard, N., Ciais, P., Viovy, N., de Noblet-Ducoudré, N., Wang, X., Magliulo, V., Wattenbach,
653 M., Vitale, L., Di Tommasi, P., Moors, E.J., Jans, W., Elbers, J., Ceschia, E., Tallec, T., Bernhofer, C.,
654 Grünwald, T., Moureaux, C., Manise, T., Ligne, A., Cellier, P., Loubet, B., Larmanou, E., Ripoche, D.,
655 2015. Orchidee-crop (v0), a new process based agro-land surface model: Model description and
656 evaluation over Europe. Geosci. Model Dev. Discuss. 8 (6), 4653-4696, DOI: 10.5194/gmdd-8-4653-
657 2015.
658 You, L., Wood, S., Wood-Sichra, U.,Wu, W., 2014. Generating global crop distribution maps: From
659 census to grid. Agric. Syst 127, 53-60, DOI: 10.1016/j.agsy.2014.01.002.

660

Figure

Click here to download Figure: Fig.1.docx

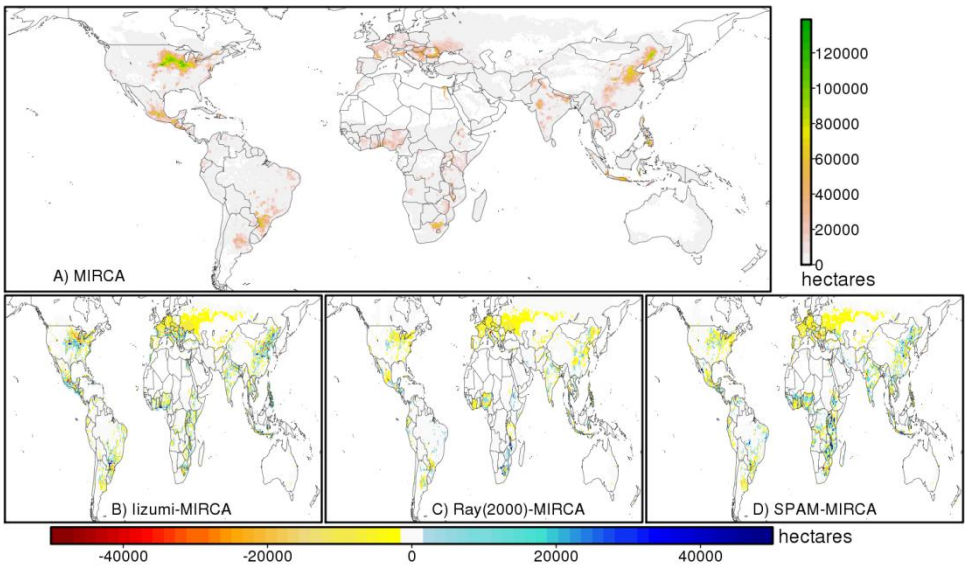

Fig. 1: Maps of spatial patterns of total harvested maize area according to MIRCA2000 (panel A) and the absolute differences in ha over 0.5° grid cells between Iizumi (panel B), Ray (time slice for the year 2000 in panel C), and SPAM2005 (panel D) and MIRCA2000 respectively. See the same figure for irrigated areas only in *SI Appendix*, Fig. B.1.

Figure

Click here to download Figure: Fig.2.docx

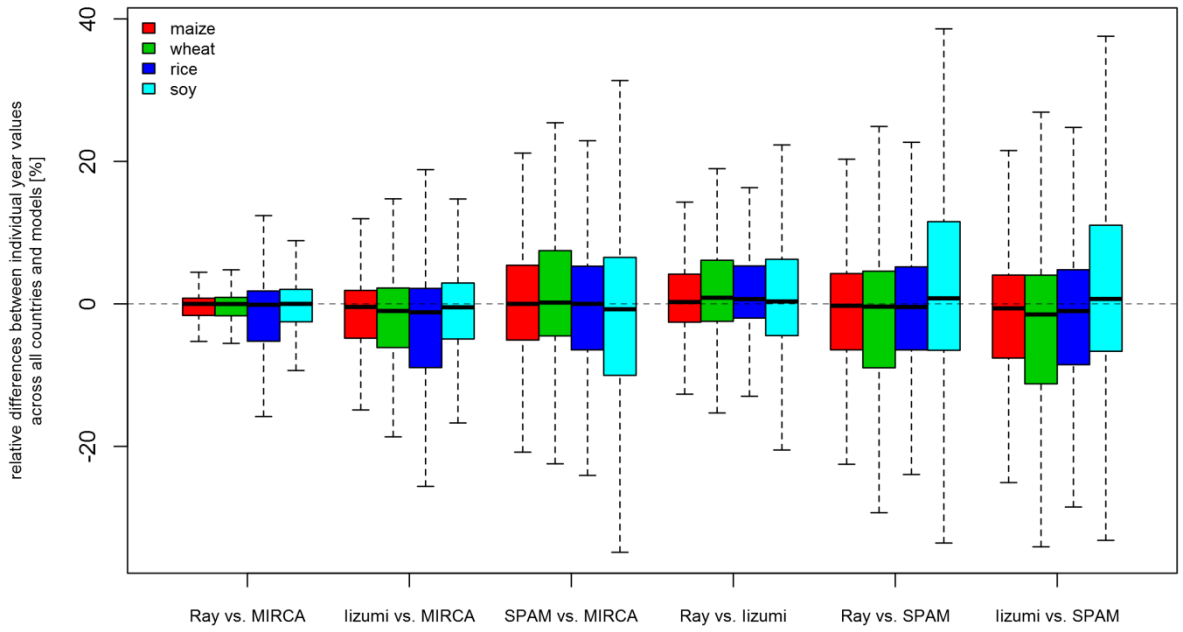

Fig.2: Boxplots of relative differences (%) between aggregated yield time series (t/ha) over 208 countries, 14 GGCMs and 31 years of the weather data set AgMERRA for the four crop types (n= 357365 for maize, n= 290061 for wheat, n= 214617 for rice, n= 202619 for soybean). Boxes show the interquartile (25-75%) range across the GGCMs used, whiskers expand to 1.5 times of inner-quartile range of national aggregated yield, and black lines within the boxes display the median value (outliers are not displayed).

Figure

Click here to download Figure: Fig.3.docx

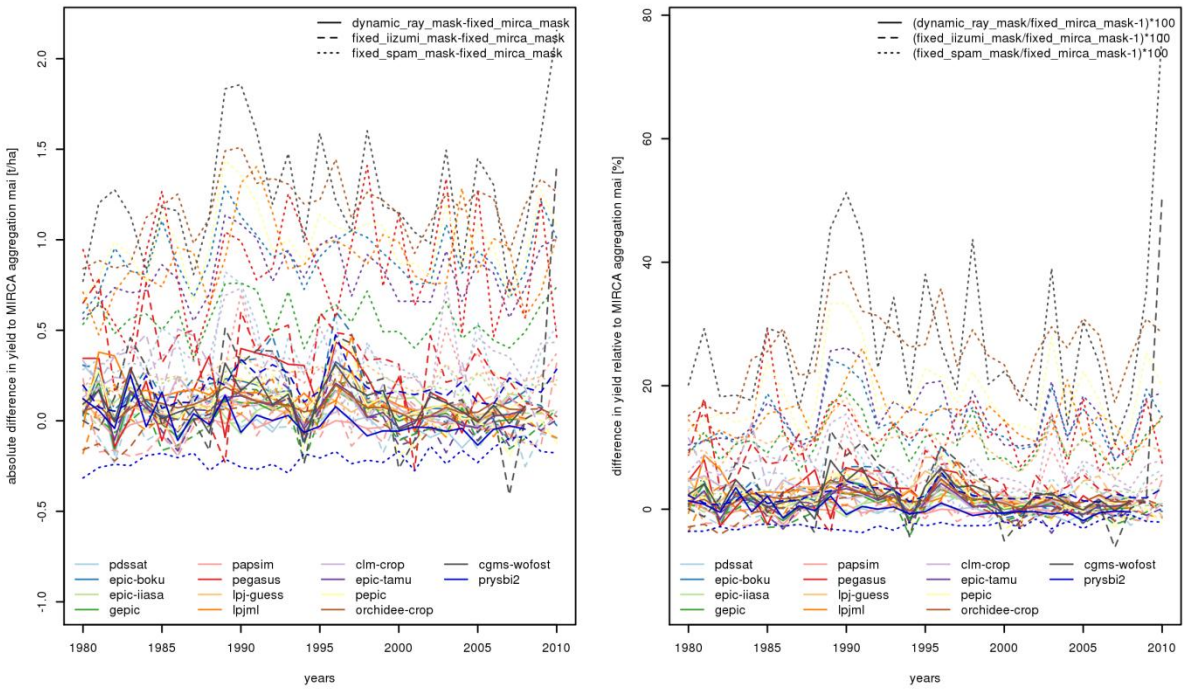

Fig. 3: Absolute (t/ha) (left panel) and (right panel) relative difference (%) between nationally annual aggregated yield time series displayed for the example case of maize yield (DM t/ha) in France. Difference per model, year, and mask from the four aggregation sets is largest with the SPAM aggregation (dotted lines) and for most models accounting for about one additional t/ha.

Figure

Click here to download Figure: Fig.4.docx

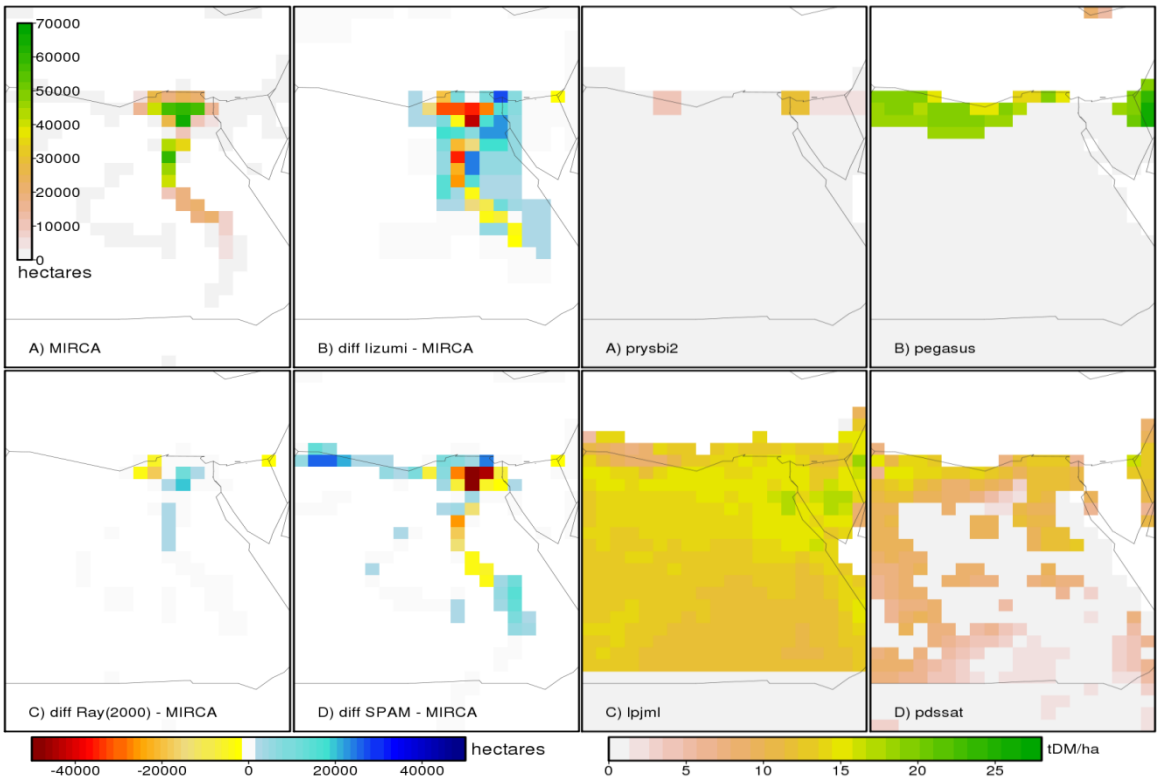

Fig. 4: (Left panel) For irrigated maize harvested areas (ha) in Egypt, spatial patterns and quantities differ between the crop masks. The maps show grid cell scale harvested area as reported by MIRCA2000 (A), and the absolute differences between harvested areas of Iizumi (B), Ray (C), and SPAM2005 (D) and MIRCA2000, respectively. (Right panel) Spatial patterns of simulated irrigated maize yields, as means over the AgMERRA weather data time period and before any masking by crop-specific harvested area data, supplied by four models A) PRYSBI2, B) PEGASUS, C) LPJmL, and D) pDSSAT. The gray shaded areas indicate grid cells where the climate conditions were regarded as unsuitable to grow irrigated maize by a model.

Figure

Click here to download Figure: Fig.5.docx

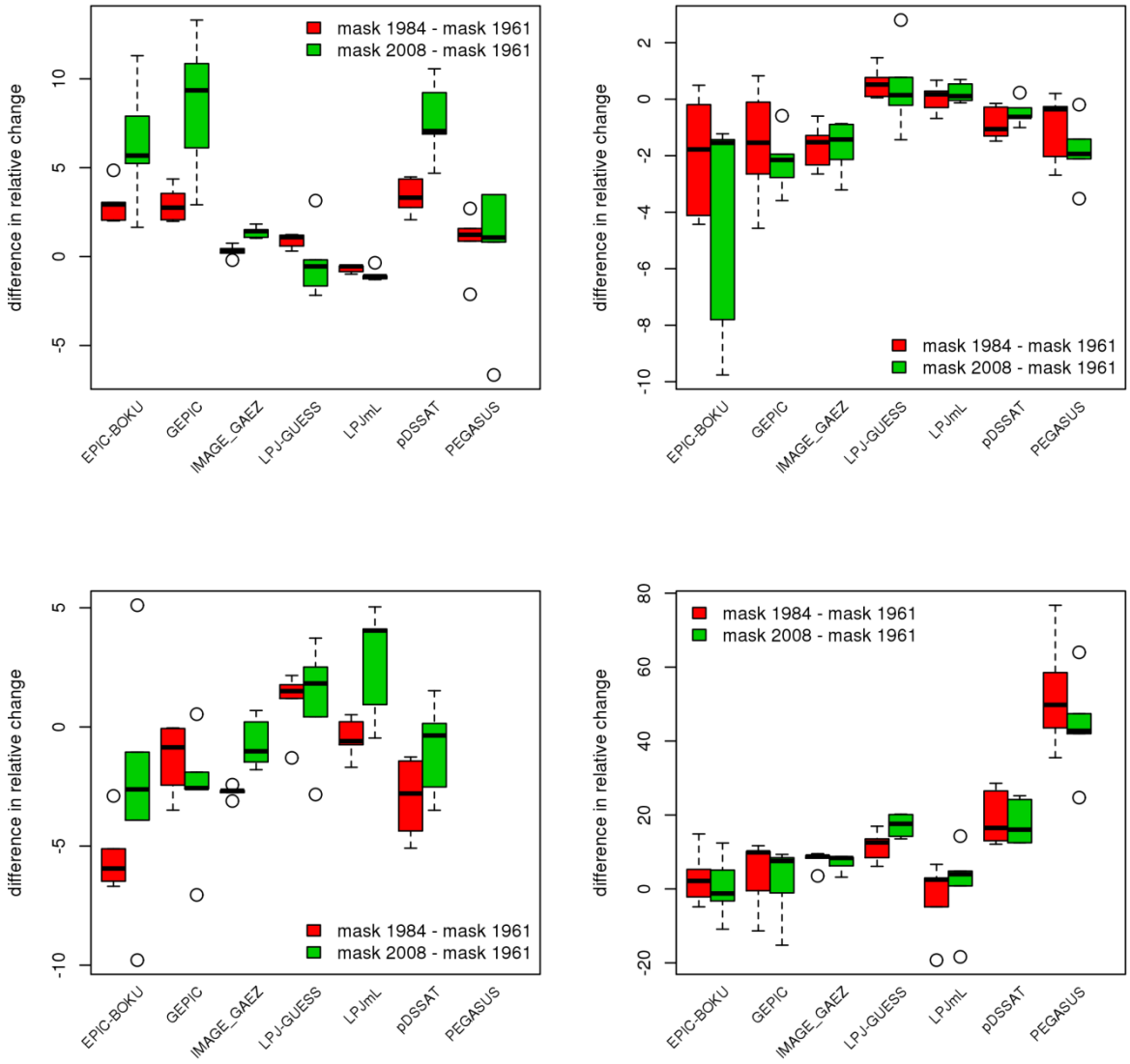

Fig. 5: Differences in projected relative yield changes (percentage change of the period 2070-2099 relative to 1980-2009) between the aggregation with the Ray crop mask of 1961, and that of 1984 (red) and 2008 (green). The panels display aggregated yields for one of the top-10 producer countries for each of the four crops: (Upper left panel) India for maize, (upper right panel) Australia for wheat, (bottom left panel) Brazil for rice, and (bottom right panel) Argentina for soybean. Boxes show the interquartile (25-75%) range across the five GCMs used, whiskers expand to 1.5 times the inner-quartile range of national aggregated yield and outliers are depicted as dots. Black lines within the boxes display the median value.

Supplementary material for on-line publication only

Click here to download Supplementary material for on-line publication only: SI_index.docx

Supplementary material for on-line publication only

Click here to download Supplementary material for on-line publication only: SI_A_Details on crop models.docx

Supplementary material for on-line publication only

Click here to download Supplementary material for on-line publication only: SI_B_Global crop specific harvested area maps.doc

Supplementary material for on-line publication only

Click here to download Supplementary material for on-line publication only: SI_C_Relative difference between data sets of total

Supplementary material for on-line publication only

Click here to download Supplementary material for on-line publication only: SI_D_G_Yield_production_harvestedarea.xlsx

Supplementary material for on-line publication only

Click here to download Supplementary material for on-line publication only: SI_E_Egypt case_graph_time_series_per_model.do

Supplementary material for on-line publication only

Click here to download Supplementary material for on-line publication only: SI_F_Aggregation of grid cell yield to FPU andcoun

Supplementary material for on-line publication only

Click here to download Supplementary material for on-line publication only: SI_References supplementary information.docx