

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/77948/>

Paper:

Ramirez-Villegas, J, Challinor, AJ, Thornton, PK and Jarvis, A (2013)
Implications of regional improvement in global climate models for agricultural impact research. Environmental Research Letters, 8 (2). ARTN 024018

<http://dx.doi.org/10.1088/1748-9326/8/2/024018>

Implications of regional improvement in global climate models for agricultural impact research

Julian Ramirez-Villegas^{1,2,3}, Andrew J Challinor^{2,3}, Philip K Thornton^{1,4}
and Andy Jarvis^{1,2}

¹ International Center for Tropical Agriculture (CIAT), Cali, Colombia

² CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), Denmark

³ Institute for Climate and Atmospheric Science (ICAS), School of Earth and Environment, University of Leeds, Leeds, UK

⁴ International Livestock Research Institute (ILRI), Nairobi, Kenya

E-mail: j.r.villegas@cgiar.org

Received 6 March 2013

Accepted for publication 22 April 2013


Published 3 May 2013

Online at stacks.iop.org/ERL/8/024018

Abstract

Global climate models (GCMs) have become increasingly important for climate change science and provide the basis for most impact studies. Since impact models are highly sensitive to input climate data, GCM skill is crucial for getting better short-, medium- and long-term outlooks for agricultural production and food security. The Coupled Model Intercomparison Project (CMIP) phase 5 ensemble is likely to underpin the majority of climate impact assessments over the next few years. We assess 24 CMIP3 and 26 CMIP5 simulations of present climate against climate observations for five tropical regions, as well as regional improvements in model skill and, through literature review, the sensitivities of impact estimates to model error. Climatological means of seasonal mean temperatures depict mean errors between 1 and 18 °C (2–130% with respect to mean), whereas seasonal precipitation and wet-day frequency depict larger errors, often offsetting observed means and variability beyond 100%. Simulated interannual climate variability in GCMs warrants particular attention, given that no single GCM matches observations in more than 30% of the areas for monthly precipitation and wet-day frequency, 50% for diurnal range and 70% for mean temperatures. We report improvements in mean climate skill of 5–15% for climatological mean temperatures, 3–5% for diurnal range and 1–2% in precipitation. At these improvement rates, we estimate that at least 5–30 years of CMIP work is required to improve regional temperature simulations and at least 30–50 years for precipitation simulations, for these to be directly input into impact models. We conclude with some recommendations for the use of CMIP5 in agricultural impact studies.

Keywords: climate model, skill, agriculture, climate change, adaptation

 Online supplementary data available from stacks.iop.org/ERL/8/024018/mmedia



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](http://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

1. Introduction

The impacts of climate change on agriculture are highly uncertain [1–3]. There is some consensus on the sign

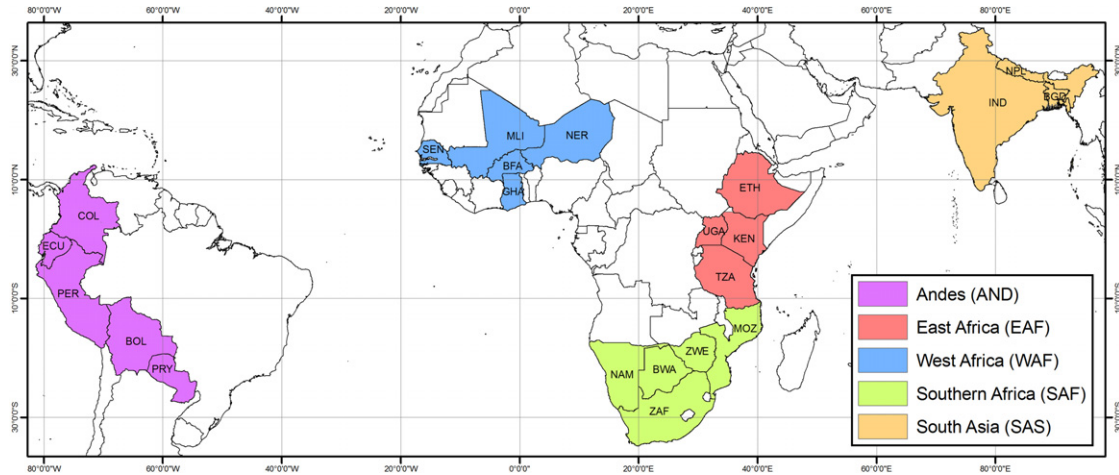


Figure 1. Regions and countries analyzed in the present study. For visualization purposes, country names were reduced to their 3-letter unique identifier (ISO), and are noted as follows: *Andes*: COL (Colombia), ECU (Ecuador), PER (Peru), BOL (Bolivia), PRY (Paraguay); *East Africa*: ETH (Ethiopia), UGA (Uganda), KEN (Kenya), TZA (Tanzania); *West Africa*: SEN (Senegal), MLI (Mali), NER (Niger), BFA (Burkina Faso), GHA (Ghana); *Southern Africa*: MOZ (Mozambique), ZWE (Zimbabwe), BWA (Botswana), NAM (Namibia), ZAF (South Africa); *South Asia*: IND (India), NPL (Nepal), BGD (Bangladesh).

of change, however, with negative effects expected on tropical annual cereals and grain legumes [1, 2, 4, 5], and positive impacts predicted for root crops [4, 6]. In assessing such impacts and any required adaptation options to abate them, future projections of climate and agricultural systems play an important role [7–9]. Nonetheless, future outlooks of agricultural production and food security are contingent on the skill of GCMs in reproducing seasonal rainfall and temperatures [10–12]. Thus, accurate climate change projections are important for developing appropriate and effective adaptation strategies and better target global emissions reduction goals. In improving projections, enhancing our understanding of important modes of variability [13, 14], the role of the different forcings in the climate system [15], as well as the responses of plants to environmental factors [16] are key steps to reducing the uncertainties that can potentially constrain adaptation [17].

The Coupled Model Intercomparison Project (CMIP) has significantly contributed to these needs, as it has coordinated nearly 20 years of climate model improvement. To date, more than 70% of impact studies carried out have used the CMIP-related model simulations (particularly those of CMIP3) as the only source of future climate projections [7, 8]. Moreover, recent estimates of global warming in CMIP models have proven to be robust [18, 19], thus enhancing our confidence on CMIP models climate projections [19]. Hence, it appears evident that the recent release of the CMIP5 (the latest phase of CMIP) ensemble [20] will likely form the basis of many future impact prediction studies. There is expectation that the increased model resolution and model complexity [20] would result in an overall reduction of model bias [21]. Nevertheless, while errors in CMIP5 GCMs have been diagnosed on a global level for key climate system features [15, 21, 22], regional assessments of model skill have been attempted for a limited number of regions [23, 24]. Importantly, as a general rule such studies are not related to agricultural impact assessment.

If impact studies that use CMIP5 are to be designed and interpreted judiciously, a critical and obvious step from the impact community is to assess the skill of impact-relevant variables in CMIP5 model simulations of historical climate. This would foster agricultural researchers’ engagement in the climate model discussion and can help agricultural researchers in deciding how to use CMIP5 model projections into impact models [25]. Furthermore, a better understanding of CMIP5 will facilitate researchers to revisit and, where necessary, make adjustments to national communications to the United Nations, national adaptation plans, scientific priority setting and research experiments aimed at informing climate change impacts and adaptation [26, 27].

In this paper, we assessed the skill⁵ [28] of 24 CMIP3 (supplementary table S1 available at stacks.iop.org/ERL/8/024018/mmedia) and 26 CMIP5 GCMs (supplementary table S2 available at stacks.iop.org/ERL/8/024018/mmedia) in five regions of the tropical world (Andes, West Africa, East Africa, Southern Africa and South Asia, figure 1)—chosen due to their vulnerability to climate change [1, 2, 4]. We assessed four variables as key climate fields exerting control on crops: mean temperature, daily temperature extremes (i.e. diurnal temperature range), precipitation, and wet-day frequency. We used four sets of observation-based data to develop robust measures of model skill: University of East Anglia Climatic Research Unit datasets [29], WorldClim [30], various sources of weather stations, and the ERA-40 reanalysis [31]. Model improvements were assessed for the five regions and compared across variables and seasons. Finally, we analyzed the implications of model errors for agricultural impact assessment, and elucidate some options for using CMIP5 data into impact models.

⁵ Skill in this study is referred to as the capacity of a climate model to represent a certain aspect of present climate (see [28]).

2. Materials and methods

Skill was assessed here using a set of metrics chosen to be in agreement with impacts and climate literature, and to be easily interpretable in both contexts. We divided skill in climate models in two different aspects: (1) mean climate, and (2) interannual variability. Although there are many measures of model skill [32], our measures are aimed at providing information that can be easily interpreted in an agricultural context. Mean climate is assessed using four metrics, and the interannual variability is assessed using only one metric.

We gathered data for precipitation, wet-day frequency, mean temperatures and diurnal temperature range for both GCMs and observations, and performed analyses for four seasons (December–January–February (DJF), March–April–May (MAM), June–July–August (JJA) and September–October–November (SON) and annual means (temperature and diurnal temperature range) and totals (precipitation, wet-day frequency).

2.1. Evaluation datasets

Here we have used various sets of observed data for all calculations. We have also used reanalysis data as it is considered to be in the upper bound of climate model skill [33]. We preferred to use various sources of observational data (of different nature) as this allows a more robust treatment of observational uncertainties. Each dataset used here would have biases and/or different degrees of suitability when an assessment of climate models is intended (e.g. station data may not be compared directly with coarse GCM grid cells, while reanalysis data for precipitation is expected to have large biases). For this reason, most studies assessing GCM predictions would attempt to use various different sets of observations, and so we have adopted a similar approach (see e.g. [23]).

For assessing mean climates, we have used four different sources of observation-based data:

- (1) CL-WST: Following [30], data were gathered from the Global Historical Climatology Network [34] (GHCN), the World Meteorological Organization Climatology Normals (WMO CLINO), FAOCLIM 2.0 (Food and Agriculture Organization of the United Nations Agro-Climatic database) [35], and a number of other minor sources (see [36]). The final dataset contained data for 35 608 locations (precipitation), 16 875 locations (mean temperature), and 12 458 locations (diurnal temperature range) at the global level. Wet-day frequency data were not available in this dataset.
- (2) CL-WCL: Global interpolated surfaces were downloaded from WorldClim [30] representative for the period 1950–2000. Global gridded data were downloaded at the resolution of 10 arcmin. Monthly maximum and minimum temperatures were used to compute diurnal temperature range. The final dataset comprised monthly climatological means of precipitation, mean temperature and diurnal temperature range for the 12 months of the year. As in

CL-WST, wet-day frequency data were not available in CL-WCL.

- (3) CL-CRU: The University of East Anglia Climatic Research Unit (CRU) high resolution interpolated climatology version 2.0 [29]. This dataset holds significant similarities to WorldClim in terms of input data, methods and gridded data [30], but is acknowledged to be more robust, as input data quality checking is reported to be much more rigorous [29]. We downloaded data at the only available resolution (10 arcmin) for monthly total precipitation, wet-day frequency, mean temperature, and diurnal temperature range.
- (4) CL-E40: Finally, the open-access version (i.e. $2.5^\circ \times 2.5^\circ$) of the European Centre for Medium-Range Weather Forecasts (ECMWF) 40+ Reanalysis (ERA-40) [31] was used as it is a fair intermediate between observations and climate model outputs [37]. We used ERA-40 as it has shown better skill than other available reanalysis products [23], and the alternative use of bias-corrected reanalysis was deemed unnecessary owing to the sometimes unexpected effects of bias correction [25], particularly in areas where moist convection is the main driver of regional precipitation as well as the similarity in temperature data in reanalysis and bias-corrected reanalysis datasets. Daily mean, maximum and minimum temperatures and total precipitation were retrieved from the ECMWF archive (at http://data-portal.ecmwf.int/data/d/era40_daily/) for the period 1961–2000. Wet-day frequency was calculated as for all other datasets as the number of days in a month with precipitation above 0.1 mm, whereas diurnal temperature range was calculated as the difference between maximum and minimum temperatures. Daily data were then aggregated to the monthly level from which mean monthly climatology was calculated. This yielded gridded ($2.5^\circ \times 2.5^\circ$) global datasets of total precipitation, wet-day frequency, mean temperature and diurnal temperature range for each month, as averages of all years in the period 1961–2000.

Data for assessing interannual variability were compiled from three sources:

- (1) TS-CRU: the CRU time series (CRU-TS3.0 at www.cru.uea.ac.uk/cru/data/hrg) of monthly precipitation, wet-day frequency, mean temperature and diurnal temperature range for the period 1961–2000 were downloaded at the resolution of 0.5° ;
- (2) TS-WST: monthly time series of precipitation, mean, maximum and minimum temperature were downloaded from the GHCN version 2 [34] dataset and were then combined with precipitation series that were previously assembled by researchers at CIAT [36]. The CIAT weather station database contained data only for precipitation, and so this was the only variable for which the two sources (i.e. GHCN and CIAT) accounted data (i.e. temperature data consisted only of GHCN stations). The final dataset, at the global level, comprised 29 736 rainfall locations

(CIAT and GHCN), 7198 mean temperature locations (only GHCN), and 4959 diurnal temperature locations (only GHCN); although not all locations had data for all months and years and many had data for less than 10 years. Wet-day frequency data were not available in TS-WST.

- (3) TS-E40: the ERA-40 daily data were again used but this time only aggregated to the monthly level for each year, which produced monthly gridded ($2.5^\circ \times 2.5^\circ$) datasets of total precipitation, wet-day frequency, mean temperature and diurnal temperature range for every year between 1961 and 2000.

All these datasets can be grouped in three categories:

- (1) point-based data (CL-WST, TS-WST), (2) gridded observed data (CL-CRU, CL-WCL, TS-CRU), and (3) gridded reanalysis data (CL-E40, TS-E40). By including these three types of data we expect to provide a broader picture of climate model skill than if only one type of data were used. Data were in all cases gridded to climate model resolution by averaging onto the model grid and these further re-gridded to $1^\circ \times 1^\circ$ for presenting all results.

2.2. Climate model simulations

GCM data were downloaded from the CMIP3 [38] and the CMIP5 archives [20]. For CMIP3, only monthly precipitation, maximum, minimum and mean temperature data were available for the 20th century simulation (i.e. 20C3M). Wet-day frequency was thus not analyzed in CMIP3. Data were downloaded for 24 GCMs (supplementary table S1 available at stacks.iop.org/ERL/8/024018/mmedia), with only three models having missing data for maximum and minimum temperatures. Time series of monthly total precipitation and means of mean monthly temperature and diurnal temperature range were computed ('TS-C3'). From these, climatological means were further calculated ('CL-C3').

In relation to CMIP3, CMIP5 has a wider range of numerical experiments [20], and data for a larger number of models and model ensembles. CMIP3 historical ('20C3M') simulations included 24 CGCMs, whereas CMIP5's included more than 35 CGCMs [20]. Here, 26 of these presented data for the historical simulation at the time queried (February 2012), although the total number of simulations is 70 (supplementary table S2 available at stacks.iop.org/ERL/8/024018/mmedia), owing to individual ensemble members. CMIP5's experimental design includes individual perturbed physics and initial conditions ensemble members for a number of models and higher resolution models (supplementary figure S1 available at stacks.iop.org/ERL/8/024018/mmedia). Similarly, models have increased their complexity by including atmospheric chemistry, aerosols, the carbon cycle, and experiments at decadal timescales [20]. Available daily outputs of historical simulations (a total of 70, supplementary table S2) for the variables of interest were downloaded and processed for the years 1961–2000 in order to produce time series of total monthly precipitation, the number of wet days, and means for mean temperature and diurnal temperature

range ('TS-C5'). Using these, climatological means for the same variables ('CL-C5') were produced. For both CMIP3 and CMIP5 ensembles multi-model means (MMM) were calculated using the climatological means and monthly time series of all GCM simulations.

2.3. Assessment of mean climates

Climate models and the MMM were assessed for their ability to represent mean climates for each of the four variables. Performance was assessed for five regions (figure 1). For each region and season we compared the climate model predictions and the observed (or reanalysis) data using all pixels in that particular geographic domain, thus assessing the time-mean geographic pattern. We used four metrics: (1) the Pearson product-moment correlation coefficient (R); (2) the root mean squared error (RMSE, equation (1)); (3) the RMSE normalized by the observed mean times 100 (RMSE_M, equation (1)); and (4) the RMSE normalized by the observed standard deviation times 100 (RMSE_{SD}).

$$RMSE_M = \frac{RMSE}{\bar{X}} = \frac{\sqrt{\frac{\sum_{i=1}^n (X_i - Y_i)^2}{n}}}{\frac{\sum_{i=1}^n X_i}{n}} \quad (1)$$

where X and Y are the observed and GCM values (respectively) for a grid cell (i) in a given season (e.g. the JJA season) over a domain with n grid cells. \bar{X} is the average of observed values. Throughout the paper, we mostly report and conclude based on the RMSE_M because (1) it was generally a better indicator of model skill as it allowed better comparisons across climate models, regions and seasons than the RMSE; (2) high values of r were not always associated with high similarity between predicted and observed values; (3) errors generally exceeded observed spatial variability by a factor of 1.5 or more, and (4) the alternative use of the standard deviation of the time series as normalizing factor was not possible due to the unavailability of these data for all mean climate datasets.

2.4. Assessment of interannual variability

Interannual variability in climate models was assessed following [32, 39], in which an interannual variability index (VI) is calculated for each climate model run as the differences between the ratios of model (M, termed TS-C3 and TS-C5 here) and observed (O, consisting of three different datasets: TS-CRU, TS-WST, and TS-E40) standard deviations (equation (2))

$$VI_v^i = \left(\frac{\sigma_{Mv}^i}{\sigma_{Ov}^i} - \frac{\sigma_{Ov}^i}{\sigma_{Mv}^i} \right)^2 \quad (2)$$

where σ is the standard deviation of the time series (1961–2000 in this study) of a variable (v) for a given grid point (i). The index is always positive and with no upper limit. As opposed to all mean climate calculations, VI calculations were performed individually for each grid cell using all years in the period 1961–2000.

2.5. Assessment of climate model improvements

In order to compare the two model ensembles, we first present some advantages of CMIP5 in terms of experimental design and model resolution. We then draw probability density functions (PDFs) using country and season values of each climate model of each ensemble. Finally, we select thresholds to classify the values of $RMSE_M$, $RMSE_{SD}$, and VI. These thresholds were then used to separate individual countries and seasons. By counting the per cent of country-season combinations above or below these thresholds out of the total (total being 22 countries times 4 seasons = 88), a more objective comparison of the differences between regions, model ensembles, and variables was achieved. For $RMSE_M$ and $RMSE_{SD}$ we chose a value of 40% (chosen to be in agreement with the value of VI chosen). For VI, we focused on the work of [39], who defined the upper bound of skilled models to be at 0.5. A value of VI below 0.5 ensures that simulated interannual variability is between -25 and $+40\%$ of the observed value. Albeit somewhat arbitrary, the thresholds chosen are consistent between them, and are likely to be representative of boundaries beyond which impact models would be severely constrained [40]. The use of other (albeit close to the present ones) thresholds showed no effect on our conclusions.

3. Results

3.1. Skill of historical climate simulations

The ability of CMIP GCMs to represent geographical variation in mean climate over a year is neither uniform nor consistent. The lowest performance is observed for precipitation in the DJF period. Conversely, seasonal mean temperatures show the highest correlations and overall lowest $RMSE_M$ and $RMSE_{SD}$ values (figure 2, supplementary figures S2–S7 available at stacks.iop.org/ERL/8/024018/mmedia). Seasonal mean temperatures in the CMIP3 model ensemble depict RMSE between 1 and 18°C ($RMSE_M$ between 2 and 130%), whereas those of CMIP5 show RMSE in the range 1 – 16°C (5–100% of the mean, 5–600% of the standard deviation) (figure 2, supplementary figures S2–S7). However, the majority of models depict $RMSE < 5^\circ\text{C}$ in both ensembles (figure 2, supplementary figures S2 and S5). Only in Nepal the majority of models show $RMSE > 10^\circ\text{C}$ in all seasons, owing to the temperature variations across the Himalayas [32]. In Africa, errors are lower, with CMIP3 model RMSE values generally between 2 and 10°C (5–40% of the mean, supplementary figure S6). CMIP5 models show lower values for all metrics. In particular for CMIP5, South African countries show RMSE values generally below 2°C (10% of the mean, supplementary figure S3). Results are consistent between observed and reanalysis datasets, hence we focus on observed datasets.

Diurnal temperature range shows higher values for all skill metrics, particularly for the Andes, where RMSE reached 15°C in the Andes (Ecuador, Peru, and Bolivia). Most models in both ensembles depict errors above 40% with respect to

the mean for all seasons and regions (figure 2, supplementary figures S2–S3), but RMSE values are very large in relation to spatial variability, and typically exceed 100%. Across Africa, night-day temperature differences are simulated more accurately as compared to Asia and the Andes (figure 2, supplementary figures S2–S7). Errors in precipitation are closely tied to errors in the wet-day frequency. Models overestimate the number of days, while underestimating the total rainfall [32]. Particularly in monsoon-driven seasonal climates such as those of West Africa and South Asia, RMSE values for seasonal precipitation and wet-day frequency (the latter only for CMIP5) are large both in absolute terms and in relation to the mean and variability (figure 2, supplementary figures S2–S7). In both ensembles, precipitation errors are the largest in South Asia (in Bangladesh RMSE values reach $2000\text{ mm season}^{-1}$), followed by the Andes, where RMSE values typically exceed $500\text{ mm season}^{-1}$ (50% of the mean) (figure 1, supplementary figures S2–S7). Seasonal precipitation RMSE is typically below $500\text{ mm season}^{-1}$ across African countries. Wet-day frequency RMSE varies between 150 and 200 days yr^{-1} .

We find interannual variability significantly misrepresented in both model ensembles. Some models, however, show strengths in some of the regions: at all grid cells, there is always at least one GCM with $VI < 0.5$, but in all cases the ensemble maxima are above this threshold. This indicated that models' skill in simulated interannual variability is generally not geographically consistent. Accurate simulations of interannual variability (i.e. $VI < 0.5$) are only achieved for mean temperature across West Africa and Southern Africa, in addition to precipitation in Southern Africa (only for CMIP5) (supplementary figure S8). The wet-day frequency, only analyzed for CMIP5, is the most poorly simulated of all variables, owing to its limited predictability [11].

3.2. Improvement in climate models

Our comparison of CMIP5 with its predecessor indicates that gains in skill have occurred mainly in simulated mean temperatures and diurnal range. We report an increase in the frequency of the left-hand side of the PDFs of $RMSE_M$. Increases are in the range 5–15% for climatological mean temperatures and 3–5% for climatological diurnal range (figure 3, supplementary figure S9). This result is consistent throughout the seasons. Conversely, gains in skill in reproducing seasonal precipitation are limited to the JJA season, with gains being limited to 1–2% increases in frequency of low $RMSE_M$ values (supplementary figure S10). Improvements in interannual variability occur to a lesser extent, but are consistent for the three variables (supplementary figures S11 and S12). Improvements in interannual variability are the largest in the JJA season, where increases in the frequency of values of VI below 0.5 are 1–5%, with little difference across variables. Importantly, the spread of the PDFs (i.e. shading in figure 3, and supplementary figures S9–S11) is lower in CMIP5 in relation to CMIP3, although the differences are not large.

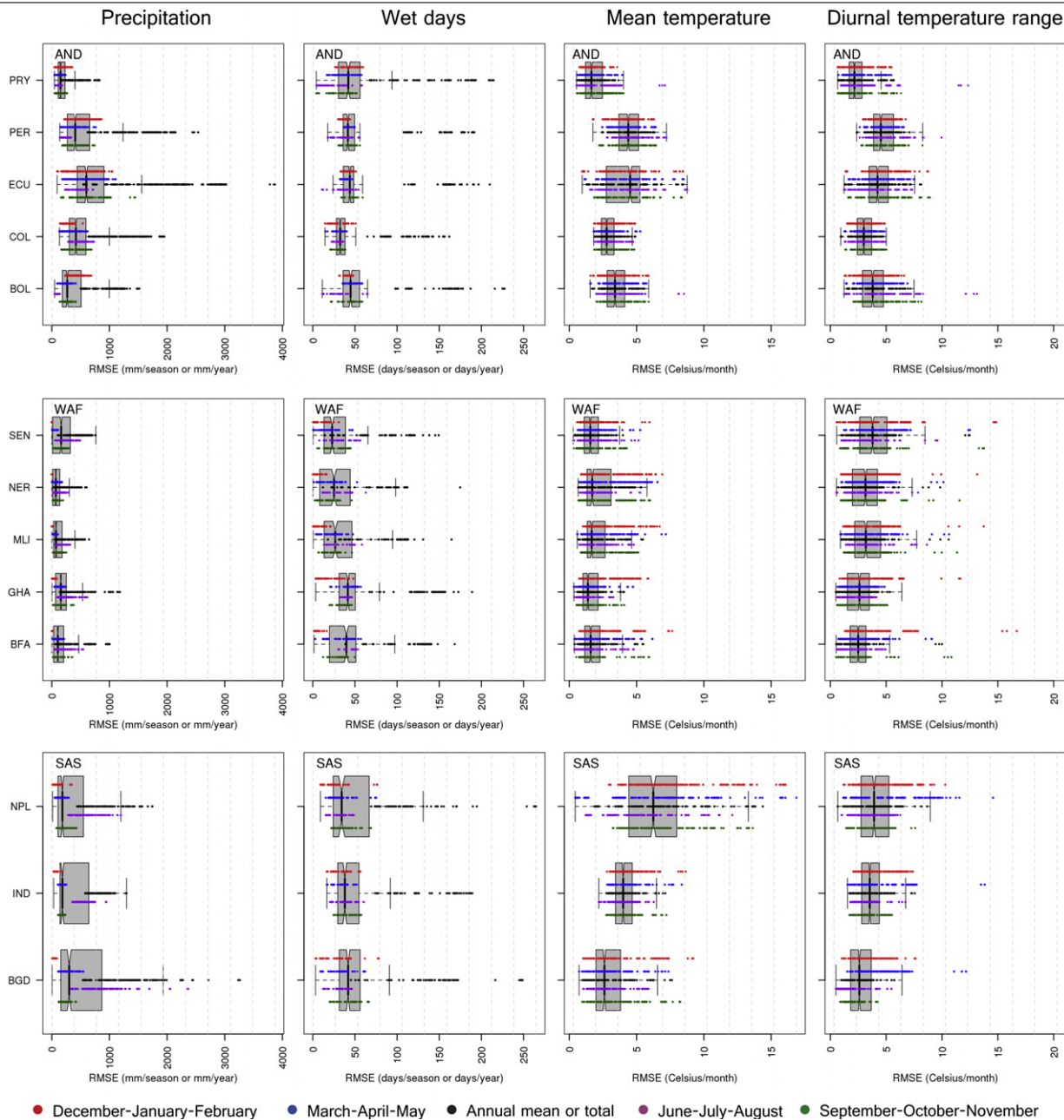


Figure 2. RMSE of all 70 CMIP5 climate model runs and the multi-model-mean across three regions and four seasons (and the annual totals or means). Plots show the distribution of all seasons and the annual total or mean for three different variables. Thick vertical lines (and notches) within the box show the median, boxes extend the interquartile range and whiskers extend 5% and 95% of the distributions. Individual points plotted along the boxes in different colors show performance for individual seasons and the annual mean (or total), as indicated by the bottom legend. Region and country typology as indicated in figure 1. See supplementary figure S2 (available at stacks.iop.org/ERL/8/024018/mmedia) for the remaining two regions.

We report significant regional differences in the increases in model skill from one ensemble to the other (figure 4, supplementary figures S12–S13). Gains in mean climate model skill are larger in South Asia for all variables. Conversely, interannual variability improved more significantly over West Africa, where skill is already relatively high (figure 4, bottom).

3.3. Implications for agricultural impact assessment

In CMIP5, model complexity has increased [19, 20]. This implies that while newly incorporated processes have

increased the physical plausibility of the models, skill has either maintained or increased (figure 3). This is likely to be an important step forward, but still only one of many needed if the final aim is to facilitate impact assessment and adaptation. Two comparisons were hereby performed to investigate the potential effect of GCM errors on agricultural impact modeling. First, we compared temperature thresholds in 12 major crops (table 1) with the mean RMSE in CMIP5’s mean climatological temperatures. We find that, if GCM outputs are used with no bias treatment into crop models, threshold exceedance could be increased by 6–69% for all

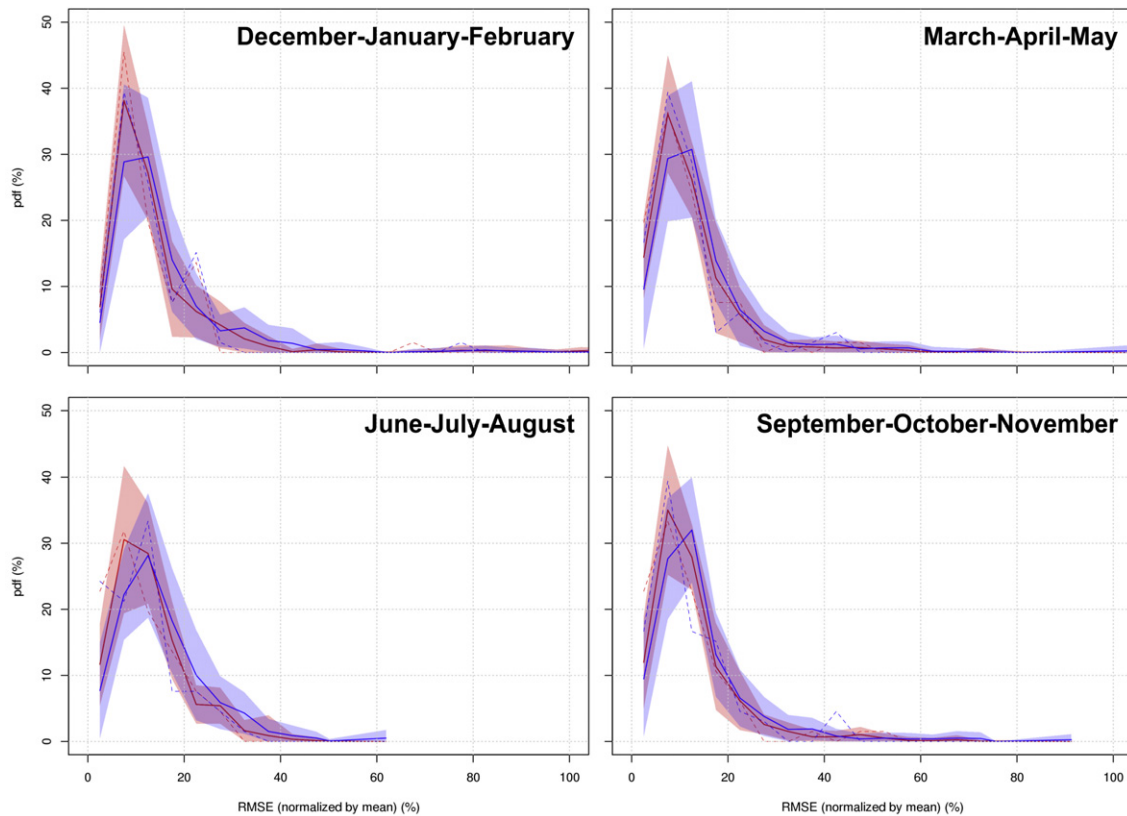


Figure 3. Improvement in skill to simulate climatological mean seasonal temperatures in CMIP5 (red) climate models with respect to CMIP3 (blue). The continuous lines show the average probability density function (PDF) of all GCMs in each ensemble and the shading shows \pm one standard deviation. Dashed lines show the PDF of the MMM. Individual GCM PDFs (including that of the MMM) were constructed using all country-season combinations of all observed datasets, except the ERA-40 reanalysis.

crops as a cause of climate model error. Although this risk has decreased by roughly 15% from CMIP3 to CMIP5 (owing to an average reduction of 15% in RMSE), this comparison suggests that raw CMIP5 simulated regional temperatures are of limited use for agricultural impact research.

Second, we used recent literature to illustrate the degree of sensitivity of cropping systems to variation in prevailing climate conditions. Although differences between future and current yields are expected to arise primarily from the climate change signal [26], model errors could bias such signal [41], hence biasing estimates of future cropping system sensitivity to climate. We find large sensitivities in cropping systems to $+2\text{ }^\circ\text{C}$ increases in mean temperature and 20% decreases in seasonal rainfall (figure 5). Median sensitivity to temperature ranges from -23% (wheat) to -12% (sorghum), whereas that related to 20% increase in precipitation was -2.5% for maize, -7.5% for sorghum and -6% for wheat. In our analyses, GCMs exceeded 20% RMSE_M in the majority of cases (supplementary figure S3) and $2\text{ }^\circ\text{C}$ in the majority of seasons, particularly in the Andes, East Africa and South Asia (figure 2).

Temperature is the primary driver of crop phenology, but also affects fertility, canopy growth, plant senescence and nutrient absorption. On the other hand, total seasonal rainfall as well as the number of days with rainfall are major drivers in the world's 50% of agricultural lands that are

rained [42]. Especially for regions where crops suffer from limited water availability, biased seasonal rainfall or wrongly timed rainfall can have large effect in model simulations [40]. Water-induced crop failures are to a large extent dependent on seasonally timed stresses that trigger death of either photosynthetic or reproductive organs in the plant. Thus, lack of soil moisture due to biased input rainfall can constrain the predictability of future crop failures (see [43]). Although the absolute effects of model errors over impact estimates are difficult to determine without crop model simulations being carried out, we argue that the use of raw CMIP5 data into impact models could significantly under- or overestimate cropping system sensitivity by 2.5–7.5% for precipitation-driven areas and 1.3–23% for temperature-driven areas.

4. Discussion and conclusions

Several questions arise from the results presented here and elsewhere [19, 32, 39]. The most overarching one, probably, is whether climate change simulations are useful for impact research. In other words, what kind of information can we usefully extract from climate models? The usefulness of climate model simulations within the context of agricultural impact research is tied to the effect of model bias on simulations of crop productivity. Literature on impacts suggests the range of information extracted from climate

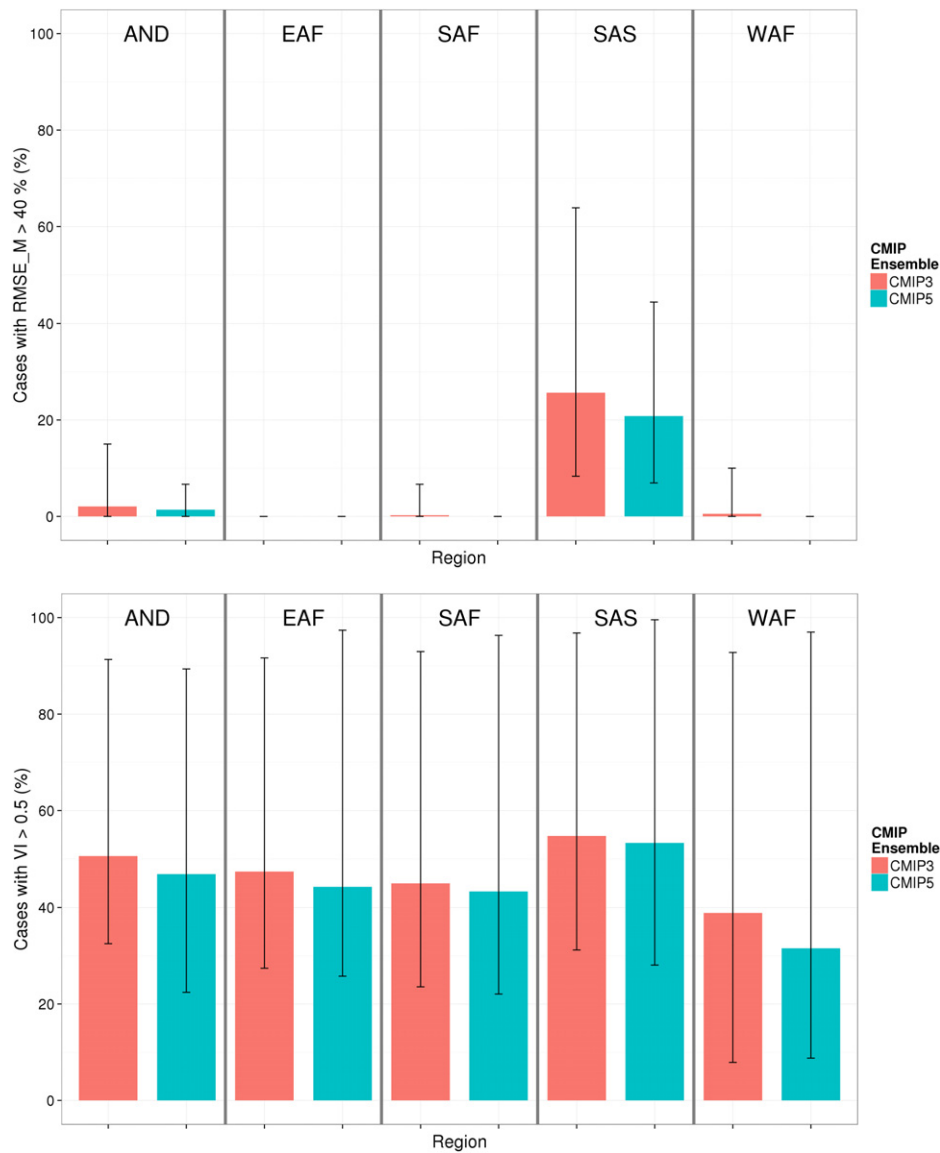


Figure 4. Regional differences in seasonal model skill and gains in model skill for seasonal mean temperatures. Top: per cent of country-season combinations where the $RMSE_M$ is above 40%. Bottom: per cent of grid cells within the analysis domain where VI is above 0.5. Bars show the average of all GCMs and error lines span the range of variation of individual GCM simulations of each ensemble. See figure 1 for region names and countries included.

models is highly varied [7, 8], ranging from the sole use of mean changes (see e.g. [44]) to the full coupling of crop-climate models (see e.g. [45]). This variation is mostly due to known and expected climate model errors (also see section 3.3).

The seasonal and regional differences in model error reported herein may also seriously hinder assessments of food systems under future scenarios, as they may imply different degrees of predictability in future impacts for crops sown at different times in the same location or for different locations sowing the same crop. In northern Indian rice-wheat systems, for instance, the large GCM biases in monsoon rainfall would make rice, sown in the rainy season, much more difficult to simulate relative to wheat, which is sown in the winter season under irrigation. Rice or wheat systems in Latin America or Africa would be differently affected, thus further complicating

assessments of global food security and, in turn, decision making. Therefore, identifying the correct pieces and amounts of information within a GCM simulation that can be used robustly into impact models is important for improving impact estimates. In that sense, a better understanding of the causes of limited model skill (see e.g. [25]) as well as of the key drivers of crop yields (see e.g. [10]) is needed. Recent research has validly focused on the ways to improve global and regional climate model simulations so as to make them useful for impact research [25, 41] and so we suggest this focus be maintained.

Identifying relevant GCM output, however, also depends on when model improvements will meet the (rather high) input standards of the agricultural research community (i.e. high accuracy at high spatio-temporal resolution). Assuming improvements have a linear trend in time we

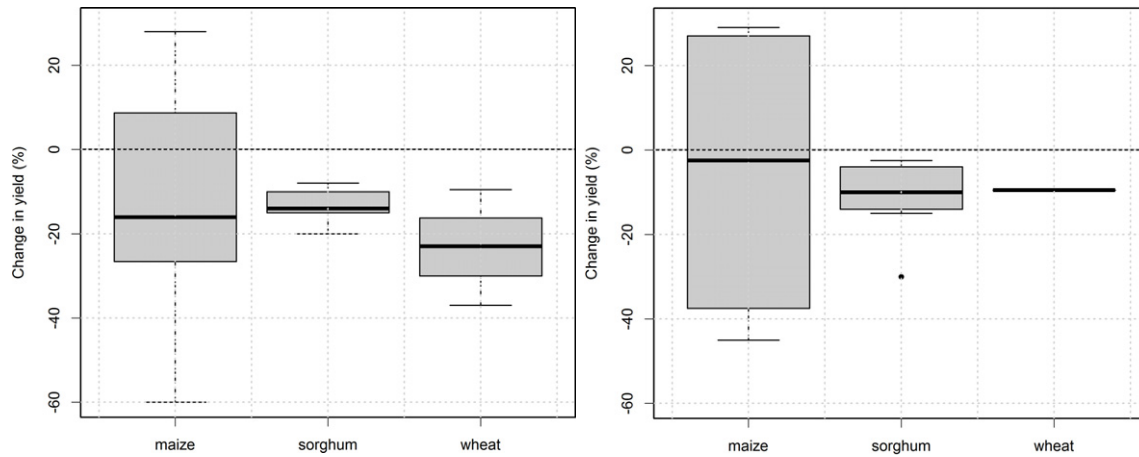


Figure 5. Sensitivity of yield in major crops to +2 °C temperature (left) and –20% precipitation (right). Boxplots have been drawn using existing literature (11 studies in total, see supplementary table S3 available at stacks.iop.org/ERL/8/024018/mmedia). As each crop is plotted using independent research articles, the boxes span spatial variability in the crop’s response. Thick vertical lines (and notches) within the box show the median, boxes extend the interquartile range and whiskers extend 5% and 95% of the distributions.

Table 1. Cardinal and extreme temperature thresholds (in °C) for major crops as reported in existing literature. Values in parenthesis correspond to the ratio of mean climatological temperature RMSE (i.e. the mean of all CMIP5 models) to the threshold specified times 100.

Crop type	Crop	Mean ^a			Extremes ^b		Reference
		T_B	T_O	T_M	T_{C1}	T_{C2}	
Cereals	Wheat	0 (Inf)	21 (13.2)	35 (7.9)	34 (8.2)	40 (7)	[52, 53]
	Maize	8 (34.8)	30 (9.3)	38 (7.3)	33 (8.4)	44 (6.3)	[54, 55]
	Rice	20 (13.9)	28 (9.9)	35 (7.9)	22 (12.6)	30 (9.3)	[56, 57]
	Barley	0 (Inf)	26 (10.7)	50 (5.6)	30 (9.3)	40 (7)	[58, 59]
	Sorghum	8 (34.8)	34 (8.2)	40 (7.0)	32 (8.7)	44 (6.3)	[55, 60]
	Millet	10 (27.8)	34 (8.2)	40 (7.0)	30 (9.3)	40 (7)	[61, 62]
Legumes	Groundnut	8 (34.8)	28 (9.9)	50 (5.6)	32 (8.7)	44 (6.3)	[63, 64]
	Soybean	7 (39.7)	32 (8.7)	45 (6.2)	30 (9.3)	44 (6.3)	[55, 65]
	Dry bean	5 (55.6)	30 (9.3)	40 (7.0)	28 (9.9)	40 (7.0)	[55, 66]
Roots and tubers	Cassava	15 (18.5)	30 (9.3)	45 (6.2)	35 (7.9)	45 (6.2)	[67, 68]
	Sweet potato	10 (27.8)	30 (9.3)	42 (6.6)	30 (9.3)	42 (6.6)	[69]
	Potato	4 (69.5)	15 (18.5)	28 (9.9)	22 (12.6)	28 (9.9)	[70, 71]

^a Cardinal temperature for development is based on a typical triangular function where T_B is the base temperature below which development ceases, T_O is the temperature at which maximum development occurs, and T_M is the maximum temperature at which development occurs.

^b Extreme temperatures are typical thresholds above which the reproductive capacity of the plant will start to be affected by high temperatures (T_{C1}) or when no reproduction will occur (T_{C2}). Values for rice correspond to night temperatures. Data for maize were unavailable and thus temperatures that would reduce (T_{C1}) or stop (T_{C2}) photosynthesis are given. For cassava and sweet potato successful flowering is irrelevant for yield, thus photosynthesis thresholds are provided instead.

estimate that at least 5–30 years of CMIP work are required to improve regional temperature simulations, while 30–50 years may be required for sufficiently accurate regional precipitation simulations, though these figures vary on a regional basis (figure 4). Ideally, the ultimate goal should be a complete coupling of crop and climate models, as this would allow an appropriate treatment of the feedbacks in the earth system [45]. Nonetheless, in ~30 years, global mean temperature would have already reached dangerous levels [46], hence stressing the need to use climate model information in an offline (i.e. not coupled), but robust and informed way. Such informed way should include: (1) an enhanced understanding of the impact of climate uncertainties

on impact estimates [26, 33], (2) improved quantification of agricultural model uncertainty, (3) a more systematic focus on the assessment of sensitivity of impact models to climate model errors [40], (4) a better quantification of downscaling and bias-correction uncertainty [47], (5) a better reporting of results by reporting impacts using raw versus downscaled and/or bias-corrected climate data [47]. Meanwhile, constant improvement in skill of predictions at higher spatio-temporal scales through more investment in climate modeling is warranted in order to meet the largely unfulfilled needs of the impact research communities [25].

A last crucial question is related to the implications of model improvements (i.e. the differences in skill between

CMIP3 and CMIP5) for impact research (also see section 3.3). This is important because as GCM ensembles increase their complexity, new research questions may arise, and also because as simulations improve in skill, impact estimates may change. It is likely that, provided enough time, CMIP5 will be widely adopted by the impact research community (see two recent examples in [48, 49]). We argue that at the very least, the replacement of CMIP3 by CMIP5 could represent an opportunity to capitalize on better climatological knowledge to identify realistic climate projection ranges [15, 50], and hence better constrain projections of crop productivity. A larger and more complex ensemble [20] could also be an opportunity to develop probabilistic (rather than deterministic) projections of climate change impacts and improve on the ways impact estimates are delivered to the public. Further research is needed, however, to understand the effects of the differences between the two ensembles on impact estimates (see e.g. [48]).

If effective and appropriate agricultural adaptation is to happen in the next 2–4 decades [51], uncertainties and lack of skill in simulated regional climates need to be communicated and understood by agricultural researchers and policy makers. One of the main barriers to adaptation lies within the skill with which climate models reproduce climate conditions. Thus, as a critical and needed step towards a better understanding of climate simulations for improving impact predictions and reducing uncertainty, we have assessed the skill of the two last CMIP model ensembles in reproducing mean climates and interannual climate variability. Further research is warranted on the diagnosis of errors in remaining impact-relevant variables (e.g. dry-spell frequency, incoming shortwave radiation, evapotranspiration, soil moisture), as well as on the effects of the differences between the two ensembles in impact estimates, as this would strengthen the conclusions reached in the present study.

Acknowledgments

This work was supported by the CGIAR-ESSP Research Program on Climate Change, Agriculture and Food Security (CCAFS). Authors thank Carlos Navarro-Racines, from the International Center for Tropical Agriculture (CIAT) for the support in downloading the CMIP5 data, and Charlotte Lau, former researcher of CCAFS, for her valuable editorial work on an earlier version of this manuscript. We thank many reviewers for their insightful comments.

We acknowledge the World Climate Research Programme's Working Group on Coupled Modeling, which is responsible for CMIP, and we thank the climate modeling groups (listed in supplementary tables S1 and S2 of this paper available at stacks.iop.org/ERL/8/024018/mmedia) for producing and making available their model output. For CMIP the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

References

- [1] Roudier P, Sultan B, Quirion P and Berg A 2011 The impact of future climate change on West African crop yields: what does the recent literature say? *Glob. Environ. Change* **21** 1073–83
- [2] Knox J, Hess T, Daccache A and Wheeler T 2012 Climate change impacts on crop productivity in Africa and South Asia *Environ. Res. Lett.* **7** 034032
- [3] Müller C, Cramer W, Hare W L and Lotze-Campen H 2011 Climate change risks for African agriculture *Proc. Natl Acad. Sci.* **108** 4313–5
- [4] Lobell D B, Burke M B, Tebaldi C, Mastrandrea M D, Falcon W P and Naylor R L 2008 Prioritizing climate change adaptation needs for food security in 2030 *Science* **319** 607–10
- [5] Thornton P K, Jones P G, Ericksen P J and Challinor A J 2011 Agriculture and food systems in sub-Saharan Africa in a 4 °C+ world *Phil. Trans. R. Soc. A* **369** 117–36
- [6] Jarvis A, Ramirez-Villegas J, Herrera Campo B and Navarro-Racines C 2012 Is Cassava the answer to African climate change adaptation? *Trop. Plant Biol.* **5** 9–29
- [7] White J W, Hoogenboom G, Kimball B A and Wall G W 2011 Methodologies for simulating impacts of climate change on crop production *Field Crops Res.* **124** 357–68
- [8] Ramirez-Villegas J and Challinor A 2012 Assessing relevant climate data for agricultural applications *Agricult. For. Meteorol.* **161** 26–45
- [9] Mearns L 2010 The drama of uncertainty *Clim. Change* **100** 77–85
- [10] Berg A, Sultan B and de Noblet-Ducoudré N 2010 What are the dominant features of rainfall leading to realistic large-scale crop yield simulations in West Africa? *Geophys. Res. Lett.* **37** L05405
- [11] Ines A V M, Hansen J W and Robertson A W 2011 Enhancing the utility of daily GCM rainfall for crop yield prediction *Int. J. Climatol.* **31** 2168–82
- [12] Lobell D B, Sibley A and Ivan Ortiz-Monasterio J 2012 Extreme heat effects on wheat senescence in India *Nature Clim. Change* **2** 186–9
- [13] Booth B B B, Dunstone N J, Halloran P R, Andrews T and Bellouin N 2012 Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability *Nature* **484** 228–32
- [14] Wang C and Picaut J 2004 Understanding ENSO physics—a review *Earth's Climate: The Ocean–Atmosphere Interaction* (Washington, DC: AGU) pp 21–48
- [15] Forster P M, Andrews T, Good P, Gregory J M, Jackson L S and Zelinka M 2013 Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models *J. Geophys. Res.: Atmos.* **118** 1139–50
- [16] Fuhrer J 2003 Agroecosystem responses to combinations of elevated CO₂, ozone, and global climate change *Agricult. Ecosys. Environ.* **97** 1–20
- [17] Wilby R L et al 2009 A review of climate risk information for adaptation and development planning *Int. J. Climatol.* **29** 1193–215
- [18] Frame D J and Stone D A 2013 Assessment of the first consensus prediction on climate change *Nature Clim. Change* **3** 357–9
- [19] Knutti R and Sedlacek J 2013 Robustness and uncertainties in the new CMIP5 climate model projections *Nature Clim. Change* **3** 369–73
- [20] Taylor K E, Stouffer R J and Meehl G A 2011 An overview of CMIP5 and the experiment design *Bull. Am. Meteorol. Soc.* **93** 485–98

- [21] Liu C, Allan R P and Huffman G J 2012 Co-variation of temperature and precipitation in CMIP5 models and satellite observations *Geophys. Res. Lett.* **39** L13803
- [22] Sillmann J, Kharin V V, Zhang X, Zwiers F W and Bronaugh D 2013 Climate extremes indices in the CMIP5 multi-model ensemble. Part 1: model evaluation in the present climate *J. Geophys. Res.: Atmos.* **118** 1716–33
- [23] Jourdain N C, Gupta A S, Taschetto A S, Ummenhofer C C, Moise A F and Ashok K 2013 The Indo-Australian monsoon and its relationship to ENSO and IOD in reanalysis data and the CMIP3/CMIP5 simulations *Clim. Dyn.* doi:10.1007/s00382-013-1676-1
- [24] Ying X U and Chong-Hai X U 2012 Preliminary assessment of simulations of climate changes over China by CMIP5 multi-models *Atmos. Ocean. Sci. Lett.* **5** 489–94
- [25] Ehret U, Zehe E, Wulfmeyer V, Warrach-Sagi K and Liebert J 2012 Should we apply bias correction to global and regional climate model data? *Hydrol. Earth Syst. Sci. Discuss.* **9** 5355–87
- [26] Challinor A J, Ewert F, Arnold S, Simelton E and Fraser E 2009 Crops and climate change: progress, trends, and challenges in simulating impacts and informing adaptation *J. Exp. Bot.* **60** 2775–89
- [27] Howden S M, Soussana J-F, Tubiello F N, Chhetri N, Dunlop M and Meinke H 2007 Adapting agriculture to climate change *Proc. Natl Acad. Sci.* **104** 19691–6
- [28] Randall D A *et al* 2007 Climate models and their evaluation *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* ed S Solomon *et al* (Cambridge: Cambridge University Press) pp 273–313
- [29] New M, Lister D, Hulme M and Makin I 2002 A high-resolution data set of surface climate over global land areas *Clim. Res.* **21** 1–25
- [30] Hijmans R J, Cameron S E, Parra J L, Jones P G and Jarvis A 2005 Very high resolution interpolated climate surfaces for global land areas *Int. J. Climatol.* **25** 1965–78
- [31] Uppala S M *et al* 2005 The ERA-40 re-analysis *Q. J. R. Meteorol. Soc.* **131** 2961–3012
- [32] Gleckler P J, Taylor K E and Doutriaux C 2008 Performance metrics for climate models *J. Geophys. Res.* **113** D06104
- [33] Iizumi T, Nishimori M and Yokozawa M 2009 Diagnostics of climate model biases in summer temperature and warm-season insolation for the simulation of regional paddy rice yield in Japan *J. Appl. Meteorol. Climatol.* **49** 574–91
- [34] Peterson T C and Vose R S 1997 An overview of the global historical climatology network temperature database *Bull. Am. Meteorol. Soc.* **78** 2837–49
- [35] FAO 2001 *FAOCLIM 2.0 A World-Wide Agroclimatic Database* (Rome: FAO)
- [36] Jones P G and Thornton P K 1999 Fitting a third-order Markov rainfall model to interpolated climate surfaces *Agric. For. Meteorol.* **97** 213–31
- [37] Challinor A J, Wheeler T R, Slingo J M, Craufurd P Q and Grimes D I F 2005 Simulation of crop yields using ERA-40: limits to skill and nonstationarity in weather–yield relationships *J. Appl. Meteorol.* **44** 516–31
- [38] Meehl G A *et al* 2007 THE WCRP CMIP3 multimodel dataset: a new era in climate change research *Bull. Am. Meteorol. Soc.* **88** 1383–94
- [39] Scherrer S C 2011 Present-day interannual variability of surface climate in CMIP3 models and its relation to future warming *Int. J. Climatol.* **31** 1518–29
- [40] Watson J and Challinor A 2013 The relative importance of rainfall, temperature and yield data for a regional-scale crop model *Agric. For. Meteorol.* **170** 47–57
- [41] Themeßl M, Gobiet A and Heinrich G 2012 Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal *Clim. Change* **112** 449–68
- [42] Mueller N D, Gerber J S, Johnston M, Ray D K, Ramankutty N and Foley J A 2012 Closing yield gaps through nutrient and water management *Nature* **490** 254–7
- [43] Baigorria G A *et al* 2010 Forecasting cotton yield in the southeastern United States using coupled global circulation models *Agron. J.* **102** 187–96
- [44] Ruane A C *et al* 2013 Climate change impact uncertainties for maize in Panama: farm information, climate projections, and yield sensitivities *Agric. For. Meteorol.* **170** 132–45
- [45] Osborne T M, Lawrence D M, Challinor A J, Slingo J M and Wheeler T R 2007 Development and assessment of a coupled crop–climate model *Glob. Change Biol.* **13** 169–83
- [46] Joshi M, Hawkins E, Sutton R, Lowe J and Frame D 2011 Projections of when temperature change will exceed 2 °C above pre-industrial levels *Nature Clim. Change* **1** 407–12
- [47] Hawkins E, Fricker T E, Challinor A J, Ferro C A T, Kit Ho C and Osborne T M 2013 Increasing influence of heat stress on French maize yields from the 1960s to the 2030s *Glob. Change Biol.* **19** 937–47
- [48] Sultan B *et al* 2013 Assessing climate change impacts on sorghum and millet yields in the Sudanian and Sahelian savannas of West Africa *Environ. Res. Lett.* **8** 014040
- [49] Hannah L *et al* 2013 Climate change, wine, and conservation *Proc. Natl Acad. Sci.* **110** 6907–12
- [50] Stott P, Good P, Jones G, Gillett N and Hawkins E 2013 The upper end of climate model temperature projections is inconsistent with past warming *Environ. Res. Lett.* **8** 014024
- [51] IPCC 2007 *IPCC Fourth Assessment Report: Climate Change 2007 (AR4)* (Geneva: IPCC)
- [52] Hunt L A and Pararajasingham S 1995 CROPSIM-WHEAT: a model describing the growth and development of wheat *Can. J. Plant Sci.* **75** 619–32
- [53] Asseng S, Foster I A N and Turner N C 2011 The impact of temperature variability on wheat yields *Glob. Change Biol.* **17** 997–1012
- [54] Tollenaar M, Daynard T B and Hunter R B 1979 Effect of temperature on rate of leaf appearance and flowering date in maize I *Crop Sci.* **19** 363–6
- [55] Jones J W *et al* 2003 The DSSAT cropping system model *Eur. J. Agron.* **18** 235–65
- [56] Peng S *et al* 2004 Rice yields decline with higher night temperature from global warming *Proc. Natl Acad. Sci. USA* **101** 9971–5
- [57] Yin X, Kropff M J and Goudriaan J 1996 Differential effects of day and night temperature on development to flowering in rice *Ann. Botany* **77** 203–13
- [58] Abiko M *et al* 2005 High-temperature induction of male sterility during barley (*Hordeum vulgare* L.) anther development is mediated by transcriptional inhibition *Sex. Plant Reprod.* **18** 91–100
- [59] Wallwork M A B, Logue S J, MacLeod L C and Jenner C F 1998 Effect of high temperature during grain filling on starch synthesis in the developing barley grain *Funct. Plant Biol.* **25** 173–81
- [60] Vara Prasad P V, Boote K J and Allen L H Jr 2006 Adverse high temperature effects on pollen viability, seed-set, seed yield and harvest index of grain-sorghum [*Sorghum bicolor* (L.) Moench] are more severe at elevated carbon dioxide due to higher tissue temperatures *Agric. For. Meteorol.* **139** 237–51

- [61] Ong C K and Monteith J L 1985 Response of pearl millet to light and temperature *Field Crops Res.* **11** 141–60
- [62] Al-Khatib K and Paulsen G M 1999 High-temperature effects on photosynthetic processes in temperate and tropical cereals *Crop Sci.* **39** 119–25
- [63] Vara Prasad P V, Boote K J, Hartwell Allen L and Thomas J M G 2003 Super-optimal temperatures are detrimental to peanut (*Arachis hypogaea* L.) reproductive processes and yield at both ambient and elevated carbon dioxide *Glob. Change Biol.* **9** 1775–87
- [64] Challinor A J, Wheeler T R, Craufurd P Q, Slingo J M and Grimes D I F 2004 Design and optimisation of a large-area process-based model for annual crops *Agric. For. Meteorol.* **124** 99–120
- [65] Boote K J, Allen L H, Vara Prasad P V and Jones J W 2011 Testing effects of climate change in crop models *Handbook of Climate Change and Agroecosystems Impacts, Adaptation and Mitigation* ed D Hillel and C Rosenzweig (London: Imperial College Press) chapter 6
- [66] Vara Prasad P V, Boote K J, Allen L H and Thomas J M G 2002 Effects of elevated temperature and carbon dioxide on seed-set and yield of kidney bean (*Phaseolus vulgaris* L.) *Glob. Change Biol.* **8** 710–21
- [67] Matthews R B and Hunt L A 1994 GUMCAS: a model describing the growth of cassava (*Manihot esculenta* L. Crantz) *Field Crops Res.* **36** 69–84
- [68] El-Sharkawy M 2012 Stress-tolerant cassava: the role of integrative ecophysiology-breeding research in crop improvement *Open J. Soil Sci.* **2** 162–86
- [69] Cen Y-P and Sage R F 2005 The regulation of rubisco activity in response to variation in temperature and atmospheric CO₂ partial pressure in sweet potato *Plant Physiol.* **139** 979–90
- [70] Marinus J and Bodlaender K B A 1975 Response of some potato varieties to temperature *Potato Res.* **18** 189–204
- [71] Streck N A, de Paula F L M, Bisognin D A, Heldwein A B and Dellai J 2007 Simulating the development of field grown potato (*Solanum tuberosum* L.) *Agric. For. Meteorol.* **142** 1–11