DR. GATIEN NOËL FALCONNIER (Orcid ID : 0000-0003-3291-650X)

DR. MYRIAM ADAM (Orcid ID : 0000-0002-8873-6762)

DR. BRUNO BASSO (Orcid ID : 0000-0003-2090-4616)

DR. ELIZABETH MEIER (Orcid ID : 0000-0003-2394-8120)

MR. TESFAYE SHIFERAW SIDA (Orcid ID : 0000-0001-6482-2669)

DR. F TAO (Orcid ID : 0000-0001-8574-0080)

DR. HEIDI WEBBER (Orcid ID : 0000-0001-8301-5424)

Article type : Primary Research Articles

Modelling climate change impacts on maize yields under low nitrogen input conditions in sub-Saharan Africa

GN Falconnier*[1], M Corbeels[1,2], KJ Boote[3], F Affholder[1], M Adam[4,5], DS MacCarthy[6], AC Ruane[7], C Nendel[8], AM Whitbread[9], E Justes[10], LR Ahuja[11], F.M. Akinseye[12], IN Alou[13], KA Amouzou[14], S.S. Anapalli[15], C Baron[16,17], B Basso[18], F Baudron[19], P Bertuzzi[20], AJ Challinor[21], Y Chen[22,23], D Deryng[24,40], ML Elsayed[25], B Faye[26], T Gaiser[26], M Galdos[21], S Gayler[27], E Gerardeaux[1], M Giner[1], B Grant[28], G Hoogenboom[3], ES Ibrahim[8], B Kamali[8], KC Kersebaum[8], SH Kim[29], M van der Laan[13], L Leroux[1,30], JI Lizaso[31], B Maestrini[18], EA Meier[32], F Mequanint[27], A Ndoli[19], CH Porter[3], E Priesack[33], D Ripoche[20], T Sida[34], U Singh[35], W Smith[28], A Srivastava[26], S Sinha[21], F Tao[22, 23, 36], PJ Thorburn[32], D Timlin[37], B Traore[38], T Twine[39], H Webber[8]

[1] AIDA, Univ Montpellier, CIRAD, Montpellier, France, [2]CIMMYT, Nairobi, Kenya, [3]University of Florida, Gainesville, FL, 32611-0570, USA, [4]CIRAD, UMR AGAP, Bobo-Dioulasso 01, Burkina Faso, [5]AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro,

Montpellier, France, [6] Soil and Irrigation Research Centre, School of Agriculture, College of Basic and Applied Science, University of Ghana, Accra, Ghana, [7]Climate Impacts Group, National Aeronautics and Space Administration Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA, [8]Leibniz Centre for Agricultural Landscape Research, Müncheberg, Germany, [9]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India, [10]PERSYST, Univ Montpellier, CIRAD, Montpellier, France, [11]USDA-ARS, Fort Collins, CO, USA, Retired, [12]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Kano, Nigeria, [13]Department of Plant and Soil Sciences, University of Pretoria, Private Bag X20 Hatfield, 0028 South Africa, [14]African Plant Nutrition Institute (APNI), West Africa Program, P.O.Box 1576, Yamoussoukro, Cote d'Ivoire, [15]USDA-ARS, Sustainable Water Management Research Unit, P.O. Box 350, Stoneville, MS 38776, United States, [16]CIRAD, UMR TETIS, F-34398 Montpellier, France, [17]TETIS, Univ Montpellier, AgroParisTech, CIRAD, CNRS, IRSTEA, Montpellier, France, [18]Department of Earth and Environmental Sciences, Michigan State University, East Lansing, MI, 48824, USA, [19]CIMMYT, Harare, Zimbabwe, [20]INRA, Agroclim, France, [21] Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK [22]Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China, [23]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China, [24]Integrative Research Institute on Transformations of Human-Environment Systems (IRI THESys), Humboldt-Universität zu Berlin, 10099, Berlin, Germany, [25]MALR-ARC, Central Laboratory for Agricultural Climate (CLAC), Giza, Egypt, [26]Crop Science Group, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn, Germany, [27]Institute of Soil Science and Land Evaluation, Biogeophysics, University of Hohenheim, Stuttgart, Germany, [28]Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, Canada, [29]University of Washington, School of Environmental and Forest Sciences, Seattle, United States, [30]CIRAD, UPR AIDA, Dakar, Senegal, [31]CEIGRAM- Universidad Politécnica de Madrid, ETSIAAB, Madrid, Spain, [32]CSIRO Agriculture and Food, Queensland Bioscience Precinct, St Lucia, Qld 4067, Australia, [33]Institute of Biochemical Plant Pathology, Helmholtz Center Munich, Neuherberg, Germany, [34]CIMMYT, Addis Ababa, Ethiopia, [35]International Center for Soil Fertility and Agricultural Development, Muscle Shoals, AL 35662, United States, [36]Natural Resources Institute Finland (Luke), Helsinki, Finland, [37]Crop Systems and Global Change Research Unit, USDA-ARS, Beltsville, MD, United States, [38]IER, Bamako, Mali, [39]University of Minnesota, Department of Soil, Water, and Climate, St. Paul, MN 55108, United States, [40]NewClimate Institute, Berlin, Germany

*gatien.falconnier@cirad.fr

Abstract

Smallholder farmers in sub-Saharan Africa (SSA) currently grow rainfed maize with limited inputs including fertilizer. Climate change may exacerbate current production constraints. Crop models can help quantify the potential impact of climate change on maize yields, but a comprehensive multi-model assessment of simulation accuracy and uncertainty in these low-input systems is currently lacking. We evaluated the impact of varying [CO_2], temperature and rainfall conditions on maize yield, for different nitrogen (N) inputs (0, 80, 160 kg N ha^{-1}) for five environments in SSA, including cool sub-humid Ethiopia, cool semi-arid Rwanda, hot sub-humid Ghana and hot semi-arid Mali and Benin using an ensemble of 25 maize models. Models were calibrated with measured grain yield, plant biomass, plant N, leaf area index, harvest index and in-season soil water content from two-year experiments in each country to assess their ability to simulate observed yield. Simulated responses to climate change factors were explored and compared between models. Calibrated models reproduced measured grain yield variations well with average rRMSE of 26%, although uncertainty in model prediction was substantial (CV = 28%). Model ensembles gave greater accuracy than any model taken at random. Nitrogen fertilization controlled the response to variations in [CO_2], temperature and rainfall. Without N fertilizer input, maize (i) benefited less from an increase in atmospheric [CO_2], (ii) was less affected by higher temperature or decreasing rainfall and (iii) was more affected by increased rainfall because N leaching was more critical. The model inter-comparison revealed that simulation of daily soil N supply and N leaching plays a crucial role in simulating climate change impacts for low-input systems. Climate change and N input interactions have strong implications for the design of robust adaptation practices across SSA, because the impact of climate change will be modified if farmers intensify maize production with more mineral fertilizer.

Keywords: *crop simulation model*, *model intercomparison*, *ensemble modelling*, *uncertainty, smallholder farming systems*.

1. Introduction

Rainfed maize production is crucial for food security and smallholder livelihoods in sub-Saharan Africa (SSA). Maize is the largest contributor to the total value of staple crop production in Western, Eastern, Central and Southern Africa (OCDE, FAO, 2016). With limited access to means of income diversification and safety nets, smallholder farmers in SSA are highly vulnerable to climate change (Connolly-Boutin and Smit, 2016; Descheemaeker et al., 2016). Temperatures are expected to increase in West, East and Southern Africa, with multi-model climate projections indicating a warming of 1 to 4°C in the decades of 2081–2100 relative to 1986–2005 depending on the Representative Concentration Pathway (RCP) considered (IPCC, 2013). Annual rainfall is expected to increase in West and East Africa (0 to +12% depending on RCP) and to decrease in Southern Africa (-5 to -10% depending on RCP) (IPCC, 2013). The impact of climate change on maize productivity across SSA is uncertain, but significant losses are expected, especially in Southern Africa (Conway et al., 2015; Lobell et al., 2008; Rosenzweig et al., 2014) and West Africa (Sultan and Gaetani, 2016). Smallholder farms in SSA usually obtain low maize yields, on average 1.8 t ha^{-1} in 2017 (FAOSTAT, 2018). These low yield levels are largely attributable to low fertilizer use, which averaged 12, 2 and 3 kg ha^{-1} for N, P and K respectively (FAOSTAT, 2018). With limited irrigation and inadequate access and use of nutrient inputs, water and nitrogen (N) stresses prevail (Folberth et al., 2013).

Process-based soil-crop models can help quantify the potential impact of climate change on maize productivity in smallholder context whilst accounting for the water and N (and/or other plant nutrient) stresses (*e.g.* Kihara et al., 2012). Soil-crop models simulate biophysical processes resulting from plant genetics, crop management, soil properties, and weather, thus tracking water, carbon, N, (phosphorous (P) to some extent) dynamics, and energy balances as plants develop through the different phenological growth phases. As such, models must consider a range of complex processes and their interactions with weather, soil, and crop management, *e.g.* the effect of soil water dynamics on nutrient supply and uptake, or the influence of soil organic matter and organic amendments on nutrient availability during the growing season. The consideration of these soil- and climate-related processes increases model complexity, number of model parameters and data demand for model calibration. Compared to simulating irrigated systems with high nutrient inputs, where water and N are less often limiting factors, the simulation of rainfed, low-input cropping systems requires more detailed model parameterization, especially of the soil processes. Model parameters related to soil water and nutrient processes are critical for the simulation of low input systems (Corbeels et al., 2016; Jones et al., 2012). Main soil processes to be taken into account are: i) soil water dynamics, including infiltration from rainfall, redistribution within the soil profile and evapotranspiration, (ii) decomposition of soil organic matter and associated mineralization of N, and (iii) N leaching below the root zone. Accurate simulation of the plant available water is crucial for simulation of crop water stress (Whitbread et al. 2017), while mineralization and leaching largely determine soil N availability for plant uptake and therefore regulate N stress on crop growth. Hence, greater uncertainty related to model processes and parameterization is expected in the responses of low-input cropping systems to climate change. For example, it is known that N stress can strongly impact crop responses to variation in [CO_2], temperature and rainfall (Affholder, 1995; Ziska et al., 1996). Furthermore, these cropping systems (which are often critical for local food security) are generally less well studied compared to the intensified mid-latitude agricultural systems that have a greater global influence (Nendel et al., 2019).

The Agricultural Model Intercomparison and Improvement Project (AgMIP) was launched in 2010 to foster increased collaboration around crop model improvement across modelling groups (Rosenzweig et al., 2013). Crop model intercomparisons have proven useful to compare consistency among models and quantify uncertainty in model predictions (Asseng et al., 2013; Bassu et al., 2014; Fleisher et al., 2017; Li et al., 2015; Ruane et al., 2017). They have reinforced the benefit of multi-model approaches, as they help identify sources of uncertainty (associated with model parameters, model structure, and model users) (Tao et al., 2018, 2020). The ensemble mean or median usually resulted as best predictors for multiple crops and for different soil and plant variables (Martre et al., 2015; Wallach et al., 2018). For example, the intercomparison of maize models (Bassu et al., 2014) allowed assessing model uncertainty in the simulated impact of climate change on maize yields under high-production conditions, *i.e.* high-input, near-potential crop growth conditions where N fertilizer inputs ranged from 60 to 255 kg N ha^{-1} and sites were irrigated or had good rainfall and thus grain yield ranged from 5 to 11 t ha^{-1}. These conditions differ considerably from the context of smallholder farmers across SSA. Bassu et al. (2014) analyzed the effect of model structure related to aboveground crop growth processes (*e.g.* simulation of net primary production of the canopy as influenced by temperature and [CO_2]) but did not deal with soil-related processes (*e.g.* N mineralization and N leaching).

Several studies relying on the calibration of a single crop model with field data, have investigated model accuracy under current climate and explored the impact of climate change on low-input smallholder systems in SSA (*e.g.* Amouzou et al., 2019; Freduah et al., 2019; Rurinda et al., 2015; Traore et al., 2017). However, the use of a single crop model precludes an analysis of simulation uncertainty related to model structure. A few studies investigated climate change and N input interactions in smallholder context with two different crop models (Faye et al., 2018b; Guan et al., 2017). Although these studies did address the issue of model uncertainty, they did not embrace the wide diversity of existing crop models. The AgMIP Global Gridded Crop Model Intercomparison study has conducted a series of model sensitivity tests to [CO_2], temperature, water, and N conditions (Franke et al., 2020), but the applied models operated on a macro-level (~0.5 degree spatial resolution) and were not calibrated against field data to capture the conditions of controlled field experiments in SSA. Thus, the accuracy and uncertainty of model simulations and model responses to the interactions between N supply and climate change in low-input systems have not been assessed for multi-model ensembles. Understanding climate change and N fertilizer input interactions will help prioritize relevant recommendations for adaptations to climate change for African smallholder farmers who currently use low levels of N inputs but will likely intensify their cropping systems with additional mineral fertilizers (Vanlauwe et al., 2014).

This study addresses three main questions, namely: (i) What is the accuracy and uncertainty of current crop model simulations of maize yield and other intermediary variables for field experiments in the context of rainfed smallholder systems in SSA? (ii) How does N fertilizer input interact with maize response to climate change (increase in [CO_2], increase in temperature, and changes in rainfall)? (iii) Does model structure (*i.e.* formalisms to account for N dynamics) and model consistency (*i.e.* the ability to accurately simulate multiple variables) explain the simulated interaction between climate change and N fertilizer input?

By doing so, we explore the hypotheses that (i) model simulations of existing maize experiments in smallholder context in SSA are more uncertain with lower accuracy than simulations of intensified cropping systems in temperate regions, (ii) crop models simulate a lower impact of [CO_2], temperature and rainfall changes in low-input (*e.g.* 0 kg N ha^{-1}) than in high-input conditions (*e.g.* 160 kg N ha^{-1}), and (iii) model structure and consistency of simulations for multiple soil and plant variables can explain diverging responses to the interaction between N inputs and climate change.

2. Materials and methods

2.1. Experimental data

We searched the literature for peer-reviewed publications in which maize field experiments under rainfed conditions were conducted during at least two cropping seasons in representative maize growing areas in SSA. The studies needed to include measurements of crop phenology (flowering and maturity dates), final grain yield and aboveground biomass at maturity, and in-season soil water dynamics for at least one growing season. Studies chosen represent a diversity of climates, soils and management conditions found across SSA for maize production. This resulted in the selection of five experimental studies that were conducted at sites respectively in Benin, Mali, Ghana, Rwanda and Ethiopia (Figure 1 and Table 1). Besides the required data on crop phenology, grain yield, aboveground plant biomass, and in-season soil water dynamics, data on in-season leaf area index (LAI) was available in at least one of the two seasons at each site except Benin. Benin and Ghana also included additional measurements of aboveground plant N accumulation during crop growth (Benin) and at maturity (Benin and Ghana). Cultivars differed across sites and were open-pollinated varieties, except in Ethiopia where a hybrid was grown. Total applied N fertilizer was 0, 64, 80, 85 and 87 kg ha^{-1} in the sites in Benin, Rwanda, Ghana, Mali and Ethiopia, respectively. There was no irrigation at any of the sites (Table 1). The experiments were extensively described, for Benin by Amouzou et al. (2018), for Mali by Traore et al. (2014), for Ghana by MacCarthy et al. (2015), for Rwanda by Ndoli et al. (2018) and for Ethiopia by

Sida et al. (2018). Soil water content to maximum rooting depth was expressed as a percentage of plant available soil water capacity (PAWC), which was calculated as the difference between the water content at the drained upper limit (DUL) and the water content at the lower extraction limit of the maize crop (LL) (both over the maximum rooting depth) (Table 1 and Table S1). The soil initial conditions (moisture and mineral N) for the simulations are given in Table S1.

To characterize each experiment regarding soil fertility, total available mineral N during the crop growing season was estimated by summing (i) measured soil mineral N prior to sowing (0-30 cm topsoil layer), (ii) N inputs from mineral fertilizer applied and (iii) N mineralized from soil organic N in the topsoil (0-30 cm) and from manure applied. Manure was applied in Mali only (Table 1). Nitrogen mineralized from soil organic matter and applied manure was estimated considering a mineralization rate of 1.5% of soil organic N per growing season, corresponding to commonly reported average mineralization rates in SSA (Bationo et al., 2007; Masvaya et al., 2017). While PAWC and the 1.5% mineralization rate were used to describe the experimental settings, this information was not forwarded to the modelling groups and they were left to address PAWC and soil N availability as per their model usual procedure.

Weather data (daily solar radiation, minimum and maximum temperatures and rainfall) for the years of the experiments were obtained from records at on-site meteorological stations at all sites. Wind speed and relative humidity for the years of the experiments were obtained from the AgMERRA climate dataset (Ruane et al., 2015). For the model simulation of the baseline climate (1980-2010), daily solar radiation, minimum and maximum temperatures and rainfall were obtained from records at the on-site meteorological stations in Benin, Mali, and Ghana and obtained from AgMERRA in Ethiopia and Rwanda. Wind speed and relative humidity were obtained from AgMERRA for the baseline climate at all sites.

2.2. Model characteristics and calibration procedure

An ensemble of 25 crop models was used for this study (Table 2 and Table S2).

These crop models present structural differences in how they model crop growth and soil processes (*e.g.* leaf area and light interception, grain yield formation, soil water dynamics, nitrate leaching, see Table 2). Of particular interest for this study was how models simulate the effect of N supply on crop growth and yield. This aspect is described in section 2.4.2.

Model simulations were executed by individual modelling groups within AgMIP (Rosenzweig et al., 2013). The model calibration entailed two phases, *i.e.* (i) partial and (ii) full calibration. For partial calibration, minimum input values required to run the model were provided, *i.e.* soil characteristics, initial soil conditions (moisture at all sites and mineral N for Benin, Mali and Ghana), crop management (sowing date, mineral and organic fertilizer inputs), weather, and observed flowering and physiological crop maturity dates (Table 1 and Table S1). In the partial calibration phase, adjustment by modelling groups to observed values was limited to setting the model parameters involved in the simulation of the time to anthesis and time to maturity. For full calibration, all measured crop and soil variables of the experiments (see section 2.1) were provided. Modelling groups could adjust the model parameters they deemed relevant to improve the model fit to observed data, using their usual methods (e.g. manual tuning or use of an optimization program). There was no knowledge sharing between the modelers and the researchers who conducted the trials during the calibration steps to guarantee that modelers from the different groups had an equal level of information on the field experiments. All sites and growing seasons were used for model calibration and no independent evaluation of simulations was performed. Each modelling group used one unique crop model. The different versions of APSIM, DSSAT and SIMPLACE-LINTUL (see Table 2) were each used by single modelling groups.

2.3. Model response to [CO_2], temperature, rainfall and N fertilizer

Responses of fully calibrated models to variation in [CO_2], temperature and rainfall were assessed, in interaction with varying mineral N input levels. Baseline years (1980-2010) were simulated with the crop management of the second growing season at each site (Table 1) for three levels of N fertilizer (0, 80, 160 kg N ha^{-1}). Response to [CO_2] was analyzed for imposed concentrations of 360 and 720 ppm. Response to temperature was assessed by increasing daily minimum and maximum temperatures by 4 °C. Response to rainfall was analyzed by multiplying baseline daily rainfall by 0.5 and 1.50. These levels represent drastic but plausible changes in environmental conditions that allow testing the sensitivity of crop models (Rosenzweig et al., 2013). A doubling of [C02] (to 720 ppm) and a +4°C temperature increase correspond to possible conditions around 2080 as predicted by climate models under RCP 8.5 (IPCC, 2013). Factorial combinations of changes in [CO_2], temperature and rainfall were not considered. For each level of [CO_2], temperature and rainfall, model simulations were run for three levels of N fertilizer (0, 80, 160 kg N ha^{-1} split in two applications during the crop growing season).

2.4. Data analysis

2.4.1. Model accuracy, uncertainty, and response to climate change factors

We analyzed model accuracy for simulated grain yield, aboveground plant biomass, maximum LAI, aboveground plant N at maturity, harvest index and in-season soil water content. Observed and simulated values were compared using the Root Mean Square Error (RMSE) and relative RMSE (rRMSE) for each of the above variables:

$$RMSE_m = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_{i,m})^2} \qquad (1)$$

$$rRMSE_m = \frac{RMSE}{\bar{O}} \times 100 \qquad (2)$$

where O_i and $P_{i,m}$ are the observed and simulated values (for model *m*) for the i^{th} measurement, n is the number of observations (*i.e.* the sum over sites, seasons, and over measurement dates per site for in-season soil water content) and $\bar{O}$ is the mean of the observed values.

To analyze uncertainty in model simulations, the coefficient of variation (CV) of the simulations with the 25 models for a given variable at a given site (both seasons) was computed as:

$$CV_s = \frac{\sigma_s}{\bar{p}_s} \times 100 \qquad (3)$$

where σ is the standard deviation of the simulated values at site *s* and $\bar{p}$ is the mean of simulated values at site *s*. CV_s was also averaged across all sites.

We assessed the value of using an ensemble of models to simulate grain yield. We started by computing the average simulated yields with ensembles of increasing number of models (n=1 to 25) for each of the ten experiments. Then we computed the relative variation between these average simulated yields and the measured yield in the experiments:

$$U_n = \frac{\sum_{i=1}^{10}|P_{ni} - O_i|}{\sum_{i=1}^{10} Oi} \times 100 \qquad (4)$$

where O_i and P_{ni} are the observed and average simulated values (for a model ensemble of size n) for the i^{th} experiment. Starting from two to 25 models, U_n was computed for a random sampling of 5% of all the $\frac{25!}{n!(25-n)!}$ combinations of models. For n=1, all combinations were evaluated.

The relative model response to a given climate change factor was computed for a particular model as:

$$R_m = \frac{P_{future,m} - P_{baseline,m}}{P_{baseline,m}} \quad (5)$$

where $P_{future,m}$ is the 31-year (1980-2010) simulated average of model *m* for the variable of interest (*e.g.* grain yield) under changed climate (altered [CO_2], temperature or rainfall, see above) and $P_{baseline,m}$ is the 31-year simulated average of model *m* for the same variable under the baseline climate (1980-2010). Here, we analyzed the relative model response to climate change for doubling [CO_2] (360 ppm to 720 ppm), temperature +4°C, 50% of baseline rainfall and 150% of baseline rainfall for N fertilizer applications of 0, 80 and 160 kg N ha^{-1}.

The relative model response to climate change R_m can take either positive or negative values. Since the coefficient of variation between models is of limited value to assess prediction uncertainty in this case, we calculated the Inter Quartile Range (IQR) of the ensemble relative to change in the simulated variable of interest (*e.g.* grain yield).

2.4.2. Model classification

We first investigated whether model structural characteristics had an influence on the model response to climate change with different N inputs. To do so, we classified the models according to (i) their capability to simulate crop responses to N inputs, and (ii) the existence of an N module with a daily time-step in the model (Table 2).

Two models (MCWLA and GLAM) did not handle crop response to N and formed the first class. Three models (PEGASUS, SARRA-H and CELSIUS) simulated responses to N input but did not include a detailed N module. These models formed the second class. In these three models, a fixed N stress factor is applied to daily biomass production. In PEGASUS, values of seasonal N stress factor were obtained by the correlation of national N fertilizer inputs and gridded yield gap fraction data (Deryng et al., 2011). In CELSIUS and SARRA-H, a seasonal N stress factor is calculated as the ratio of total seasonal available N to the crop N uptake required for non-limited growth. In CELSIUS, total seasonal available N is calculated with mineralization coefficients obtained from the literature (Ricome et al., 2017). In SARRA-H, the N stress factor was calibrated with on-farm and on-station experiments across West Africa.

Twenty models handled crop responses to N and had a detailed N module with daily time-step calculations of soil and plant N processes; they formed the third class of models (Table 2). All class 3 models use as inputs (i) soil mineral N content at initiation of the simulation, and (ii) the amounts of fertilizer N applied at specified dates during the cropping season. These models include the explicit representation of a number of organic C pools in the soil (Table 2) and functional processes of organic matter mineralization to compute the availability of mineral N for crop uptake. In these models, daily mineralization of organic nitrogen is simulated with one to seven organic carbon and nitrogen pools (Table 2) with specific decomposition rates. Simple approaches usually identify a labile (fast decomposition rate) and a stable (slow decomposition rate) organic matter pool. More complex models have additional microbial biomass-related pools to simulate the role of soil organisms in the N mineralization-immobilization turnover process during decomposition.

Within this third class of models, we investigated whether model consistency, *i.e.* model ability to adequately simulate different soil and plant variables, could explain model performances and model responses to climate change and its interaction with N fertilizer inputs. The indicator used for model consistency was the sum of ranks (Martre et al., 2015) for rRMSE over the variables of interest (*i.e.* grain yield, total aboveground biomass, maximum LAI, total aboveground plant N, harvest index and soil water contents). Models below the median sum of ranks for rRMSE over all the variables were classified as “most consistent” models (class 3a), models above the median as “less consistent” models (class 3b) (Table 2). An alternative ranking of models was computed based on the sum of ranks for grain yield and total aboveground biomass only (the two variables available for all experimental situations). Models below the median sum of ranks for rRMSE over these two variables were classified as “highest ranked” models (for grain and biomass) (Table 2).

The effect of model class on the model response to climate change (doubling [CO_2], temperature +4°C, 50% of baseline rainfall and 150% of baseline rainfall) was examined using linear mixed model regression analysis with model class (3a or 3b) and N input as fixed factors and site as a random factor. *P*-values to test the significance of model class were obtained by likelihood ratio tests of the full regression model (including all fixed and random factors) against a regression model with only N input and site effects. Visual inspections of residuals plots did not reveal deviations from normality or heteroscedasticity. The analysis was done using R (R Development Core Team, 2019; http://www.R-project.org, last accessed 13/07/2019) and the linear mixed-effect model was coded and tested with the R package *lme4* (http://cran.r-project.org/web/packages/lme4/index.html, last accessed 16/07/2019). We performed the likelihood ratio test with the *anova* function.

3. Results

3.1 Characterization of sites and crop experiments

Seasonal rainfall (from maize sowing to harvest) varied greatly across sites and seasons, from 217 mm (Rwanda, 2014 season) to 923 mm (Ethiopia, 2014 season) (Table 1). Seasonal rainfall was low in Rwanda in 2014 but residual soil water at sowing was substantial (i.e. 57% of PAWC). Crop water stress occurred during the two experimental years in Rwanda (Figure 1B). In Benin, Mali and Ethiopia, observed soil water contents never went below 50% of PAWC during crop growth in the experiments where soil water was monitored (Figure 1B), indicating a likely low occurrence of crop water stress. In Ghana, water content was monitored to 30 cm soil depth only, so these data were of limited value for analyzing water stress. Overall, observed maize grain yields were not correlated to seasonal rainfall (Figure S1), confirming the role of N (Figure 1C) and other crop growth limiting factors in determining grain yield.

Estimated total available mineral N during the crop growing season varied widely across sites (Figure 1C). It was lowest at the experimental site in Benin, where there was no fertilizer input (Table 1). Total available mineral N was highest in the experiments in Rwanda, due to fertilizer inputs and a high soil organic N content compared with the experiments in the other sites (Table 1). Although maize yield tended to increase with estimated total mineral N availability (see section 2.1), the correlation was not significant (Figure 1C).

3.2 Model simulations of the experiments

3.2.1 Model accuracy

When partially calibrated to phenology only, most models failed to accurately reproduce grain yield variations across sites and experiments (Figure 2A); rRMSE averaged across models for grain yield was 63% (Figure 2B). Full calibration greatly improved the models' ability to reproduce observed grain yields (Figure 2A); rRMSE averaged across models decreased to 26% (Figure 2B). The median of the fully calibrated model ensemble closely approximated observed grain yields (Figure 2A). Improvement in model accuracy with full calibration was also important for aboveground biomass at maturity and maximum LAI but was more limited for aboveground plant N at maturity and harvest index (Figure 2B). Maize phenology was accurately simulated by the fully calibrated models, with rRMSEs of 8 and 13% for the sowing-anthesis and anthesis-maturity durations respectively. Regarding the temporal dynamics, the range of simulated values of in-season LAI, soil water content, aboveground plant biomass and aboveground plant N mostly enveloped the observed values (Figure S3). With exception of the 2013 season in Ethiopia, most models were able to reproduce seasonal soil water dynamics, a crucial variable for simulating crop growth when water stress occurs. The increase in soil water up to field capacity during (i) the vegetative crop phase in the field experiment in Benin in 2015 and (ii) during the reproductive phase in the field experiment in Mali in 2010 was well reproduced by most models. The decrease in soil water below 50% of PAWC early in the season in 2014 and later in the season in 2015 in Rwanda was also well simulated by most models. Main disagreements between model simulations and field measurements occurred (i) in Rwanda in 2015, for which most models underestimated LAI and overestimated aboveground plant biomass and (ii) in Ethiopia in 2013, for which all models underestimated observed aboveground plant biomass and soil water. The latter may, however, be due to errors in rainfall recording or poor calibration of the moisture probes used to estimate soil water.

Nitrogen mineralized from soil organic matter and N leached below the root zone were not measured in the field experiments so we could not assess model prediction accuracy for these variables. The ensemble median of simulated N mineralization, averaged over the two crop growing seasons, was 22, 20, 39, 43 and 38 kg ha^{-1} in Benin, Mali, Ghana, Rwanda and Ethiopia, respectively. These simulated values matched reasonably well with the empirical estimates of N mineralization using a rate of 1.5% of soil organic N (see section 3.1), *i.e.* 16, 10, 32, 93 and 40 kg ha^{-1} in Benin, Mali, Ghana, Rwanda and Ethiopia, respectively. The ensemble median of simulated N leaching, averaged over the two crop growing seasons, was 11, 15, 2, 2 and 4 kg ha^{-1} in Benin, Mali, Ghana, Rwanda and Ethiopia, respectively.

3.2.2 Model prediction uncertainty

Full model calibration resulted in a reduction of prediction uncertainty (expressed as CV), and this reduction was larger for grain yield and aboveground plant biomass at maturity than for the other plant-related variables (maximum LAI, aboveground plant N at maturity and harvest index) (Figure 2C). Overall, there was no clear indication that model prediction uncertainty was largest in the most constrained (N-limiting) sites (*e.g.* Benin, see Figure S2). Prediction uncertainty was relatively low for maize phenology (full calibration), with a CV of 9% for the sowing-anthesis duration, and 16% for the anthesis-maturity duration. Prediction uncertainty of simulated N mineralization was large, both with partial (CV of 90%) and full calibration (CV of 85%). A similar behavior was found for simulated N leaching, with CVs of 171 and 136% with partial and full calibration, respectively.

The average absolute difference between measured and simulated grain yield decreased rapidly with the number of models considered in an ensemble (Figure 3). A least eight calibrated models were needed to fall below a 13.5% threshold, *i.e.* the CV of measured yield typically obtained in experimental plots (Taylor et al., 1999).

3.2.3 Model classification

Models of class 1 and 2 simulated grain yield accurately with rRMSE values equal to or below 18% (Table 3). Some models of these classes also performed well for the other variables (*i.e.* total aboveground biomass at maturity, maximum LAI, harvest index and soil water) with rRMSE values close to or below 30%.

The ten "most consistent" models of class 3, *i.e.* models below the median sum of rank for rRMSE across all variables (Figure S4) were grouped in class 3a, and the others were placed in class 3b (Table 2). The most consistent crop model (DNDC) when considering all variables had a sum of rank of 32 (Table 3). Decrease in model uncertainty from partial to full calibration for simulated grain yield was similar for both model classes 3a and 3b, *i.e.* 57 and 42% for class 3a and 3b respectively. However, the decrease in model uncertainty for aboveground plant N at maturity was greater for models of class 3a than 3b, *i.e.* 44 and 11%, respectively, indicating a likely greater effect of calibration on N supply and N uptake for models of class 3a than 3b. After full calibration, class 3a models had a significantly ($P < 0.05$) smaller RMSE for grain yield, aboveground plant biomass at maturity, aboveground plant N at maturity, maximum LAI, harvest index and in-season soil water content compared with class 3b models. Most of the modelling groups (60%) who used class 3a models reported calibration of soil parameters related to the size of the different soil organic matter pools to adjust the amount of N mineralized from soil organic matter and to improve the match with observed aboveground plant N, while only 10% of the class 3b modelling groups reported such parameterization procedure (Table S3). Similarly, the majority (60%) of the class 3a modelling groups reported calibration of parameters related to soil water dynamics (*e.g.* moisture contents at field capacity and wilting point, soil water evaporation coefficients) to mimic observed soil water dynamics, while only 30% of the class 3b modelling groups reported such parameterization procedure (Table S3). Classifying class 3 crop models according to grain yield and aboveground biomass only (*i.e.* the variables that were observed for all sites and experiments) led to minor changes in the classification; the eight 'most consistent' models were also among the eight best models when ranked based on grain yield and aboveground biomass only (see underlined models in Table 2).

3.3 Model ensemble response to climate change and N inputs

Across sites and levels of N fertilization, the model ensemble median indicated a 4% increase in grain yield for doubling [CO_2], 21% decrease with increasing temperature (+4°C), 1% decrease with increasing rainfall (150% of baseline rainfall) and 17% decrease with decreasing rainfall (50% of baseline rainfall). Nitrogen fertilizer input controlled to a large extent the response to variation in [CO_2], temperature and rainfall (Figure 4). We describe the interactions between N fertilizer input levels and climate change factors in the subsections below.

3.3.1 Variations in [CO_2] and temperature interact with N inputs

The impact of increased [CO_2] on maize grain yield was smaller when N was limiting (Figure 4). With doubling [CO_2], the model ensemble median for the grain yield response was smaller with 0 kg N ha^{-1} (4% across all sites, *i.e.* 0.04 t ha^{-1}) than with 160 kg N ha^{-1} (7% across all sites, *i.e.* 0.29 t ha^{-1}). Model response varied across the sites (Table S4) and ranged between 0 and 5% for 0 kg N ha^{-1}, and between 4 and 13% at 160 kg N ha^{-1}.

Without N fertilization maize grain yield was less affected by higher temperature (+4°C) compared with N fertilization (80, 160 kg N ha^{-1}) (Figure 4). Across all sites, the ensemble median indicated a 14 and 26% decrease in grain yield as a result of increased temperature with 0 and 160 kg N ha^{-1}, respectively. The negative effect of higher temperature was stronger at the warm sites (Benin, Mali and Ghana) than at the cool sites (Rwanda and Ethiopia). With 160 kg N ha^{-1}, maize grain yield decreased by 29% in Benin, 32% in Ghana and 39% in Mali, and by only 14% in Ethiopia and 16% in Rwanda (Table S4).

Prediction uncertainty, expressed here as the IQR of ensemble relative response in simulated maize yield, was greater for temperature than for [CO_2] variation, without a clear indication that uncertainty decreases with increasing N fertilizer inputs (Figure S5).

3.3.2 Variation in rainfall in interaction with N inputs

Comparing the effect of N fertilization under conditions of increased rainfall (150% of baseline), grain yields of the 0 N treatment were more negatively affected than those with inputs of 80 or 160 kg N ha^{-1} (Figure 4). Across all sites, the model ensemble median indicated a -8 and 0% change in grain yield caused by increased rainfall at 0 and 160 kg N ha^{-1}, respectively. In Ethiopia, Mali, and Benin, an increase in rainfall had a strong negative effect on grain yield, and the magnitude of this effect was stronger for low N conditions. The ensemble median indicated a 7% decrease in Mali, a 16% decrease in Ethiopia and a 35% decrease in Benin at 0 kg N ha^{-1}, and 0, -4 and -2% in those countries at 160 kg N ha^{-1} (Table S4). In Ghana, and Rwanda, increased rainfall had little effect on grain yield when no N was applied (-2% and +1% relative yield change respectively) while positive effects of increased rainfall occurred with 80 and 160 kg N ha^{-1} (6 and 20% yield increase respectively).

Without N fertilization maize grain yield was less affected by a decrease in rainfall (50% of current) than with N fertilization (80, 160 kg N ha^{-1}) (Figure 4). Across all sites, the model ensemble median indicated a 2% and 27% decrease in grain yield with 0 and 160 kg N ha^{-1}, respectively. Model response varied across the sites (Table S4). The impact of a decrease in rainfall was lower for Ethiopia and Benin (20 and 4% yield decrease at 160 kg N ha^{-1}, Table S4) than for Mali, Ghana and Rwanda (25, 36 and 50% yield decrease at 160 kg N ha^{-1}, Table S4), which is consistent with the fact that Ethiopia and Benin had higher seasonal rainfall (Table 1).

Prediction uncertainty, expressed here as IQR of ensemble relative response in simulated maize yield, for rainfall variation was always higher at low input (0 kg N ha^{-1}) than at high N input (80, 160 kg N ha^{-1}) with the exception of Mali for 50% of the baseline rainfall (Figure S5). Decrease in model prediction uncertainty from low to high N input simulations was generally greater for 150% relative rainfall than for 50% decrease in rainfall (Figure S5).

3.3.3 Impact of model classification on model response to climate change in interaction with N inputs

Classifying the crop models (Table 2) allowed unravelling some of the variability related to the interaction between climate change and N fertilizer inputs. Two models, MCWLA and GLAM (class 1, Table 2), do not simulate responses to N inputs, and hence the interaction between climate change and N input could not be analyzed (Figure 5). Three models (PEGASUS, SARRA-H, and CELSIUS, see Table 2) simulate a response to N input but do not include a detailed N module. These three models had different responses to climate change and N input compared with the ensemble model responses described in sections 3.3.1 and 3.3.2. The simulated response by the SARRA-H model to increased [CO_2] was higher under the zero N fertilization than under the 80 and 160 kg N ha^{-1} fertilization in Mali. The PEGASUS and CELSIUS models simulated very little interaction between increase in [CO_2] and N fertilization. Similarly, the simulated impact of increased temperature (+4°C) by SARRA-H was largest with the zero N fertilization in Rwanda, Ethiopia and Benin, *i.e.* the opposite of the simulated trend by the model ensemble (see section 3.3.1). The PEGASUS and CELSIUS models simulated also very little interaction between increase in temperature and N fertilization. The three models (SARRA-H, PEGASUS and CELSIUS) simulated no interaction between 150% of the baseline rainfall and N fertilization. The SARRA-H and PEGASUS models simulated little to no interaction between 50% of the baseline rainfall and N fertilization, while CELSIUS predicted an interaction consistent with the model ensemble behavior. The response averaged across these three models is shown in Figure 5 (class 2 models).

The magnitude of model responses to some climate change factors was different between class 3a (the ten most consistent models ranked using all the measured variables) and the "less consistent" class 3b models (Figure 5). Simulated impact of doubling [CO_2] was significantly lower (P < 0.05) for models of class 3a than for those of class 3b. The class 3a models predicted a 0.9 and 5.3% increase in grain yield with doubling [CO_2] at 0 and 160 kg N ha^{-1}, respectively, while the class 3b models predicted a 4.0 and 11.8% increase in grain yield. On the other hand, simulated responses to changes in temperature and rainfall did not differ significantly between class 3a and 3b models (Figure 5). When ranked based on grain yield and aboveground biomass only (Table 2), highest ranked models did not differ significantly in their response to [CO_2] and rainfall. The simulated response to increased temperature (+4°C) was however significantly lower (P < 0.05) for highest ranked class 3a models (considering grain and aboveground biomass) than for the lower ranked class 3b models.

Models of class 3 simulated N leaching, whereas models of the other classes did not. This resulted in a stronger negative impact of increased rainfall on simulated grain yield, especially for zero N fertilization, *i.e.* class 3 models simulated an increase in N leaching with an increase in rainfall (Figure S6). The simulated increase in the amount of N leached with 150% of baseline rainfall did not differ significantly between the model classes 3a and 3b. Models of class 3 explicitly simulated N mineralization unlike the models of the other classes. They, however, did not simulate an increase in N mineralization when temperature was increased (Figure S7).

4 Discussion

Low input systems and model accuracy and uncertainty

Our comparative analysis of model accuracy with partial and full calibration confirms the importance of calibration against observed harvest and in-season variables for accurate simulation of maize growth and yield in smallholder context, as was the case in other model intercomparisons (e.g. Asseng et al., 2013; Bassu et al., 2014; Li et al., 2015). However, rRMSE for grain yield averaged over the fully calibrated models was greater (rRMSE = 26%) for the cropping situations in our study with relatively low inputs than for *e.g.* high-input situations in a wheat model intercomparison (rRMSE ~ 10%) (Asseng et al., 2013). This confirms our initial hypothesis that model simulations are less accurate for low-input and below potential yield situations where soil processes need to be adequately simulated. Model ensembles gave greater accuracy than any model taken at random; in our study an ensemble of at least eight randomly-selected models was needed to fall below the typical 13.5% variation of measured grain yields in field experiments. This number is in line with the findings of the previous maize, rice and wheat model intercomparison studies (Asseng et al., 2013; Bassu et al., 2014; Li et al., 2015), and demonstrates the strength of model ensembles. Model ensembles combine models that have complementary strengths in simulated plant and/or soil processes and minimize errors in structure/parameterization that may exist for some processes in individual models.

Model calibration for soil processes appeared to be key for low-input systems. For example, a steep decrease in soil water content occurred during the growing season in the experiments in Rwanda, the site with the lowest seasonal rainfall, and most models were generally able to capture such behavior. Notably, modeling groups who reported the calibration of specific parameters related to soil water dynamics to match observed soil water, achieved a greater increase in accuracy from partial to full calibration (see section 3.2.3). A correct simulation of soil mineralization was crucial for accurately simulating maize growth and yield in Benin as no N fertilizer was applied. However, the lack of observations precluded the analysis of model accuracy for N mineralization. The uncertainty in simulated N mineralization was large and not reduced with full calibration, though some models did calibrate the sizes of the organic matter pools and achieved a more accurate simulation of maize N uptake (see section 3.2.3). As expected, models simulated higher N leaching in the wetter sites (Ethiopia and Benin), but without observations we could not analyze model accuracy with respect to amounts of N leached. The large uncertainty in simulated N leaching was not reduced with full calibration, and only one model reported changes in parameter values related to N leaching with full calibration.

Our model classification indicated that the 'most consistent' models (class 3a) (see section 3.2.3) achieved a greater reduction in RMSE for aboveground plant N after full calibration, hence increasing the likeliness of obtaining accurate simulations that consistently describe the plant growth processes leading to grain yield (Martre et al., 2015). Eventually, the 'most consistent' models simulated grain yield better, *i.e.* with a significantly smaller RMSE compared with 'less consistent' models. Good calibration can however be impeded by data availability, *e.g.* aboveground plant N was not measured in Mali, Ethiopia and Rwanda. Due to imbalance in data availability between sites, modelers made assumptions on some inputs and/or model parameters, leading to uncontrolled uncertainty in model simulations. When detailed data on soil is limited, simple models with a limited number of parameters should have an advantage over more complex models (Castañeda-Vera et al., 2015). Our findings partly supported this argument. Class 3a models all used a simple "tipping bucket" model approach for water dynamics, suggesting that the more detailed Richards equation for the flow of water in unsaturated soils was not needed to simulate water stress in a satisfactory manner. However, simple models with only one single pool for the simulation of organic matter decomposition and associated N mineralization were not systematically among the 'most consistent' models.

Data quality can also impede good model calibration (Kersebaum et al., 2015), *e.g.* disagreement between (all) model simulations and soil water measurement in Ethiopia in 2013 points to issues with regard to rainfall input data, and/or soil water measurements, and/or errors in the soil textural properties leading to higher predicted water percolation through the soil profile.

Low-input cropping systems and climate change impacts

Our study revealed substantial interactions between N input and the effect of climate change. With a doubling in [CO_2], the model ensemble median for relative grain yield response was +7% across all sites at 160 kg N ha^{-1} but only +4% at 0 kg N ha^{-1}. Such simulated increase at high N fertilizer input is consistent with the previous maize model intercomparison study that indicated a 7.5% yield increase with doubling[CO_2] (Bassu et al., 2014). The range of impacts depending on sites was however narrower for our study, *i.e.* 4-13% compared with 0-19% in Bassu et al. (2014), indicating possible improvements of some models that were used in both studies (Table 2). In addition to a very small yet controversial direct effect of [CO_2] on C_4 crops photosynthesis (Leakey et al., 2006; Ziska et al., 1999), maize benefits from elevated [CO_2] because of a "water-saving" effect (taken into account in the majority of the models, see Table 2) due to reduced stomatal conductance and plant transpiration (Durand et al., 2018). The associated increase in plant growth as a result of this effect requires greater rates of N uptake and assimilation by the plant (Bunce, 2014; Stitt and Krapp, 1999). The maize model ensemble simulated smaller gains from elevated [CO_2] at 0 kg N ha^{-1} than at higher rates (160 kg N ha^{-1}), because the beneficial effects of elevated [CO_2] were constrained by N stress when no fertilizer was applied (Figure 4). Chamber-based and free-air CO_2 enrichment experiments for maize were most often conducted under optimal nutrient supply in temperate climates (Allen et al., 2011; Chun et al., 2011; Manderscheid et al., 2014). An exception is the study of Bunce (2014) that showed lower maize yield response to elevated [CO_2] as N fertilization decreased, in line with our model estimates of the impact of elevated [CO_2] for different N fertilization levels.

When N availability was limiting plant growth under 0 kg N ha^{-1}, maize models simulated only minimal impact of higher temperature and reduced rainfall, *i.e.* N stress made climate stresses less prominent. These model results are (i) supported by experimental data showing that crops with low supply of nutrients are less exposed to water stress (Affholder, 1995; Rötter et al., 1997) and (ii) in line with other modelling studies showing a less negative impact of climate variability and change in cropping systems with lower inputs (Affholder, 1997; Faye et al., 2018b; Rurinda et al., 2015; Sultan et al., 2014; Traore et al., 2017). The "Liebig law of the minimum" helps understand such pattern: growth is dictated not by total resources available, but by the scarcest resource (limiting factor). Besides, low nutrient supply causes lower leaf area index and, therefore, less transpiration compared with crops grown under non-limiting nutrient supply, leading to a lower soil water uptake by the crop and consequently less impact of drought stress when rainfall becomes insufficient (Affholder, 1997; Faye et al., 2018b).

An increase in average temperature impacts maize grain yield mainly through a reduced duration of the crop cycle and associated lower biomass accumulation and thus N uptake, a process well accounted for in current maize models (Bassu et al., 2014). We could not find any experimental work studying possible effects of N supply on crop growth duration. The lower impact of temperature under low N fertilizer input was not due to an increase in N mineralization and soil N availability: the models did not simulate increased N mineralization under increased temperature (see section 3.3.3 and Figure S7). Although higher temperatures are known to lead to an increase in N mineralization (Guntiñas et al., 2012), in the model simulations a decrease in topsoil moisture may occur as a result of increased soil water evaporation with increased temperature, thus reducing N mineralization rate and offsetting the increased mineralization due to the rise in temperature.

Maize was more affected by a projected increase in rainfall when N was limiting (0 kg ha^{-1}). We attribute this effect to the simulated increase in N leaching with increased rainfall, in line with another modelling report in a smallholder context in West Africa (Freduah et al., 2019). Simulated increase in N leaching with increased rainfall is supported by field experimental studies on tropical soils in Eastern and Southern Africa, that observed highest N leaching in growing seasons with highest rainfall amounts (Kamukondiwa and Bergström, 1994; Mapanda et al., 2012; Russo et al., 2017).

Overall, the site influenced the impact of climate change. Maize growth and yield in the cooler high altitude sites, *i.e.* Rwanda and Ethiopia, were less affected by increase in temperature, in line with other studies predicting smaller crop yield losses, and in some situations even gains at cooler locations (Waha et al., 2013; Bassu et al., 2014; Zhao et al., 2017). At low N fertilizer inputs, maize at the site with the highest level of soil organic carbon (i.e. Rwanda, see Table 1) was less affected by an increase in rainfall and the associated N leaching, highlighting the crucial role of soil organic matter in the steady provision of N in low-input cropping systems (*e.g.* Wood et al., 2018). Maize yield at sites with higher seasonal rainfall (*i.e.* Benin and Ethiopia) was less affected by the simulated decrease in rainfall, highlighting the importance of current climate conditions when analyzing the impact of climate change (Waha et al., 2013).

We found no evidence that model uncertainty regarding the response to elevated [CO_2] and temperature would be greater at low levels of N input. However, uncertainty of model response to rainfall change decreased (except in Mali) with the level of N fertilization, indicating that models differed in the way they dealt with this interaction. The high variability in simulated soil N mineralization (see section 3.2.2) explains to an extent such uncertainties.

Influence of model structure on simulated crop responses to climate change

Our analysis of crop model response to climate change coupled with experimental work suggests that accurately accounting for both N supply and N leaching under different experimental conditions is crucial for modelling climate change impacts on maize growth in SSA. By separating models into classes, we disentangled some of the variability in model response to climate change under contrasting N fertilizer inputs. Most models without a daily N module (models of class 2) did not account for the interactions in the case of increased [CO_2] and change in rainfall in a way that was consistent with experimental evidence (see section 3.3 and discussion above). Class 3a models (ranked based on rRMSE for all the observed variables) simulated a smaller impact of elevated [CO_2] on maize yield irrespective of the N input levels. There were, however, no obvious structural model characteristics differentiating these best models from the others. For light utilization, models using a "radiation use efficiency" approach or a "gross photosynthesis – respiration" approach were represented equally within the two classes. Similarly, models with specific formalisms to compute grain number were represented in the two classes. Class 3a models also simulated more accurately crop response to N input than the other models (see section 3.2.3); therefore, their simulation of the impact of climate change with contrasting N inputs is expected to be more robust. Ranking models based on various plant and soil variables may however be disputable since each variable has a different degree of importance for modelling crop growth. For this reason, we investigated an alternative ranking based on grain and biomass yield only. With this approach, the highest ranked models simulated significantly less impact of an increase in temperature irrespective of the N fertilizer level. There were, however, no obvious model structural characteristics differentiating these highest ranked models from the others, *e.g.* for the type of heat stress simulated or the formalism for crop phenology. It should, however, be noted that uncertainty in calibration due to model user subjectivity can sometimes hide the role of specific model structures (Confalonieri et al., 2016). For example, the PHINT parameter (interval between successive leaf tip appearances) in the DSSAT model can be optimized to improve accuracy in LAI and grain yield simulations (Table S3). Whether such optimization without detailed leaf appearance data to calibrate against is a good practice is a point of debate. Identifying highest ranked models prior to simulation is challenging: a given model will often obtain a different ranking for fit to the observations when used with a different dataset (*i.e.* another combination of physical environment and management) (Wallach et al., 2018). Without model validation with independent datasets (*e.g.* Confalonieri et al., 2009), it is unlikely that the ranking proposed in this study holds for all possible environments in a smallholder context. The ranking should therefore be seen as a means to understand model behavior rather than a prescription on which model to use. Eventually, in some cases model response may have been unrealistic, *e.g.* relative grain yield change with doubling [CO_2] between 50% and more than 100% (*i.e.* outliers not shown in Figure 4). Systematically discarding models with such unrealistic behavior could help in model selection and improve ensemble model creation. However, such procedure remains in dispute as discarding 'extreme' models can lead to overconfidence in models that behave in a similar way, rewarding a convergence that may be the result of similar model assumptions and errors (Knutti, 2010). Analysis of unrealistic behavior relying on relative changes also deserve caution, as very small values with baseline climate can cause very large relative responses with future climate even if the absolute responses are reasonable.

Implications for sustainable intensification in SSA

A substantial proportion of the farm households in SSA face food insecurity (Frelat et al., 2016). Sustainable intensification with increased nutrient inputs and efficient use could drastically increase crop production and improve household food availability, whilst maintaining other important ecosystem services and preventing further land expansion (Loon et al., 2019; Vanlauwe et al., 2014). Our modelling study indicates that farmers intensifying maize production will face a different impact of climate change. With increased N fertilization maize will benefit more from elevated [CO_2], but will be increasingly negatively impacted as temperature increases and/or if rainfall decreases. The benefits from elevated [CO_2] in mitigating drought impacts are unlikely to offset negative impacts from changes in temperature and possibly rainfall (*e.g.* Faye et al., 2018b), so that yield penalties and larger yield variability are expected. Increased yield variability may exacerbate the current risk of unfavorable benefit-cost ratio for mineral fertilizer application (*e.g.* Bielders and Gérard, 2015; Falconnier et al., 2017). Policy interventions aiming at implementing risk coping mechanisms and additional safety nets will therefore be crucial to support sustainable intensification in the context of climate change.

Our findings have implications for developing recommendation domains for specific adaptation strategies. In high rainfall sites like in Ethiopia and Benin, nitrate leaching will be further intensified in case of a wetter climate; technologies maximizing N efficiency and preventing losses through leaching, *e.g.* relay intercropping with deep rooting cover crops and split applications of mineral fertilizer, may prove successful. In low rainfall sites like the site in Rwanda, maize will experience more severe drought stress if climate gets drier; drought tolerant cultivars and water-harvesting technologies (e.g. stone lines, tied ridging, zaï pits and contour ridging) may help mitigate production losses. Low altitude warm sites (like in Ghana, Mali and Benin) will be more affected by the rise in temperature so that breeding should aim at cultivars adapted to heat stress.

Avenues to extend the work

Given the importance of accurately accounting for N dynamics when modelling the response of low-input systems to climate change, further model improvement studies targeting these systems should focus on (i) the evaluation of model ability to accurately simulate soil organic matter mineralization, soil mineral N dynamics (*e.g.* leaching), plant N uptake and N stress effects on crop growth by comparing simulations with observed data, and (ii) studying the impact of model structure and complexity (*e.g.* 'tipping bucket vs Richards equation, number of soil carbon pools, impact of temperature and moisture on soil organic matter mineralization) on the accuracy of model outputs. Comprehensive datasets to perform such analysis currently do not exist for SSA. The research agenda on modelling the effects of climate in low-input conditions should therefore aim at implementing detailed soil-crop monitoring in experiments in contrasting sites representative of SSA. An experimental focus on the interaction between N fertilization and elevated [CO_2] and temperature will also be required, as models have not been tested against experimental data coming from tropical environments for these interactions. Model sensitivity to rainfall was assessed in this study by assuming a uniform relative change in daily rainfall throughout the growing season. More complex patterns are likely to occur in the future, *e.g.* increase in the frequency and magnitude of intense rainfall events (Taylor et al., 2017), or shortening of the rainy season (Guan et al., 2017). More analyses of model responses that account for these complex patterns are required. Most soils across SSA are highly weathered and inherently poor in P (Buerkert et al., 2001). In three of the five experimental study sites (*i.e.* Mali, Ghana and Rwanda), substantial amounts of P fertilizer (~25 kg P ha^{-1}) were applied, which is considered as sufficient to reach about 70% of the water-limited yield potential (ten Berge et al., 2019; Velde et al., 2014). With such amount of P fertilizer, it is unlikely that P stress was an issue in these sites. For the other sites, accounting for P stress may help to reduce model uncertainty. The number of models able to deal with P stress is however limited (*e.g.* Dzotsi et al., 2010). Although maize is the most important staple food crop in large parts of SSA, other traditional cereals such as sorghum and millet are also widely consumed in West and East Africa (OCDE and FAO, 2016). Other crops such as cassava and banana also contribute substantially to food security in sub-humid and humid SSA (OCDE and FAO, 2016). Extending model intercomparisons of climate change impact for these other crops that are often cultivated in environments different from the ones of our study sites would therefore allow for a more comprehensive assessment of diverse smallholder farming systems and food security issues. Besides, climate change is likely to strengthen pest and disease pressure on crops (Deutsch et al., 2018). Although the soil-crop models used in this study do not account for biotic stresses, considering this yield-reducing factor will be a necessary step towards a more integrated assessment of the impact of climate change (*e.g.* Donatelli et al., 2017) on smallholder farming systems.

Conclusion

Our modelling study revealed robust simulated interactions between climate change factors and N fertilization and indicates that maize intensively managed with more N fertilizer will be more sensitive to climate change. Therefore, the needed sustainable intensification of cropping systems in SSA will become more and more risky as climate changes, which highlights the need for policy interventions aiming at implementing risk coping mechanisms. Predicting the impact of climate change on cropping systems in which N inputs are likely to vary, requires crop models that explicitly account for N stress and N leaching. At least eight fully calibrated models were needed to ensure reasonable accuracy in simulations. Experimental data and model improvements are urgently needed to better evaluate the impact of the interaction between (i) N fertilization and elevated [CO_2] and (ii) N mineralization and elevated temperature. We advocate for a research agenda geared towards filling the current data gap by implementing detailed and comprehensive soil-crop monitoring in contrasting sites representative of agriculture in SSA.

Acknowledgments:

We are grateful to the members of the AgMIP leadership team for their support and to Senthold Asseng and Pierre Martre for sharing their insights in the AgMIP Wheat Team. The lead author also thanks Sonali McDermid for her help in extracting AgMERRA data.

Data sharing and data accessibility

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References:

Affholder, F., 1997. Empirically modelling the interaction between intensification and climatic risk in semiarid regions. Field Crops Research 52, 79–93. https://doi.org/10.1016/S0378-4290(96)03453-3

Affholder, F., 1995. Effect of organic matter input on the water balance and yield of millet under tropical dryland condition. Field Crops Research 41, 109–121. https://doi.org/10.1016/0378-4290(94)00115-S

Allen, L.H., Kakani, V.G., Vu, J.C.V., Boote, K.J., 2011. Elevated CO2 increases water use efficiency by sustaining photosynthesis of water-limited maize and sorghum. Journal of Plant Physiology 168, 1909–1918. https://doi.org/10.1016/j.jplph.2011.05.005

Amouzou, K.A., Lamers, J.P.A., Naab, J.B., Borgemeister, C., Vlek, P.L.G., Becker, M., 2019. Climate change impact on water- and nitrogen-use efficiencies and yields of maize and sorghum in the northern Benin dry savanna, West Africa. Field Crops Research 235, 104–117. https://doi.org/10.1016/j.fcr.2019.02.021

Amouzou, K.A., Naab, J.B., Lamers, J.P.A., Becker, M., 2018. CERES-Maize and CERES-Sorghum for modeling growth, nitrogen and phosphorus uptake, and soil moisture dynamics in the dry savanna of West Africa. Field Crops Research 217, 134–149. https://doi.org/10.1016/j.fcr.2017.12.017

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rötter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P.K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainty in simulating wheat yields under climate change. Nature Clim. Change 3, 827–832. https://doi.org/10.1038/nclimate1916

Baron, C., Benjamin, S., Maud, B., Benoit, S., Seydou, T., Lebel Thierry, Janicot Serge, Dingkuhn Michael, 2005. From GCM grid cell to agricultural plot: scale issues affecting modelling of climate impact. Philosophical Transactions of the Royal Society B: Biological Sciences 360, 2095–2108. https://doi.org/10.1098/rstb.2005.1741

Basso, B., Cammarano, D., Troccoli, A., Chen, D., Ritchie, J.T., 2010. Long-term wheat response to nitrogen in a rainfed Mediterranean environment: Field data and simulation analysis. European Journal of Agronomy 33, 132–138. https://doi.org/10.1016/j.eja.2010.04.004

Bassu, S., Brisson, N., Durand, J.-L., Boote, K., Lizaso, J., Jones, J.W., Rosenzweig, C., Ruane, A.C., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., De Sanctis, G., Gayler, S., Grassini, P., Hatfield, J., Hoek, S., Izaurralde, C., Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Kim, S.-H., Kumar, N.S., Makowski, D., Müller, C., Nendel, C., Priesack, E., Pravia, M.V., Sau, F., Shcherbak, I., Tao, F., Teixeira, E., Timlin, D., Waha, K., 2014. How do various maize crop models vary in their responses to climate change factors? Glob Change Biol 20, 2301–2320. https://doi.org/10.1111/gcb.12520

Bationo, A., Kihara, J., Vanlauwe, B., Waswa, B., Kimetu, J., 2007. Soil organic carbon dynamics, functions and management in West African agro-ecosystems. Agricultural Systems 94, 13–25. https://doi.org/10.1016/j.agsy.2005.08.011

Bielders, C.L., Gérard, B., 2015. Millet response to microdose fertilization in south–western Niger: Effect of antecedent fertility management and environmental factors. Field Crops Research 171, 165–175. https://doi.org/10.1016/j.fcr.2014.10.008

Biernath, C., Gayler, S., Bittner, S., Klein, C., Högy, P., Fangmeier, A., Priesack, E., 2011. Evaluating the ability of four crop models to predict different environmental impacts on spring wheat grown in open-top chambers. European Journal of Agronomy 35, 71–82. https://doi.org/10.1016/j.eja.2011.04.001

Brisson, N., Ruget, F., Gate, P., Lorgeou, J., Nicoullaud, B., Tayot, X., Plenet, D., Jeuffroy, M.-H., Bouthier, A., Ripoche, D., Mary, B., Justes, E., 2002. STICS: a generic model for simulating crops and their water and nitrogen balances. II. Model validation for wheat and maize. Agronomie 22, 69–92. https://doi.org/10.1051/agro:2001005

Buerkert, A., Bationo, A., Piepho, H.-P., 2001. Efficient phosphorus application strategies for increased crop production in sub-Saharan West Africa. Field Crops Research 72, 1–15. https://doi.org/10.1016/S0378-4290(01)00166-6

Bunce, J.A., 2014. Corn Growth Response to Elevated CO₂ Varies with the Amount of Nitrogen Applied. American Journal of Plant Sciences 05, 306–312. https://doi.org/10.4236/ajps.2014.53042

Castañeda-Vera, A., Leffelaar, P.A., Álvaro-Fuentes, J., Cantero-Martínez, C., Mínguez, M.I., 2015. Selecting crop models for decision making in wheat insurance. European Journal of Agronomy 68, 97–116. https://doi.org/10.1016/j.eja.2015.04.008

Challinor, A.J., Wheeler, T.R., Craufurd, P.Q., Slingo, J.M., Grimes, D.I.F., 2004. Design and optimisation of a large-area process-based model for annual crops. Agricultural and Forest Meteorology 124, 99–120.

Chun, J.A., Wang, Q., Timlin, D., Fleisher, D., Reddy, V.R., 2011. Effect of elevated carbon dioxide and water stress on gas exchange and water use efficiency in corn. Agricultural and Forest Meteorology 151, 378–384. https://doi.org/10.1016/j.agrformet.2010.11.015

Confalonieri, R., Acutis, M., Bellocchi, G., Donatelli, M., 2009. Multi-metric evaluation of the models WARM, CropSyst, and WOFOST for rice. Ecological Modelling 220, 1395–1410. https://doi.org/10.1016/j.ecolmodel.2009.02.017

Confalonieri, R., Orlando, F., Paleari, L., Stella, T., Gilardelli, C., Movedi, E., Pagani, V., Cappelli, G., Vertemara, A., Alberti, L., Alberti, P., Atanassiu, S., Bonaiti, M., Cappelletti, G., Ceruti, M., Confalonieri, A., Corgatelli, G., Corti, P., Dell'Oro, M., Ghidoni, A., Lamarta, A., Maghini, A., Mambretti, M., Manchia, A., Massoni, G., Mutti, P., Pariani, S., Pasini, D., Pesenti, A., Pizzamiglio, G., Ravasio, A., Rea, A., Santorsola, D., Serafini, G., Slavazza, M., Acutis, M., 2016.

Uncertainty in crop model predictions: What is the role of users? Environmental Modelling & Software 81, 165–173. https://doi.org/10.1016/j.envsoft.2016.04.009

Connolly-Boutin, L., Smit, B., 2016. Climate change, food security, and livelihoods in sub-Saharan Africa. Reg Environ Change 16, 385–399. https://doi.org/10.1007/s10113-015-0761-x

Conway, D., Garderen, E.A. van, Deryng, D., Dorling, S., Krueger, T., Landman, W., Lankford, B., Lebek, K., Osborn, T., Ringler, C., Thurlow, J., Zhu, T., Dalin, C., 2015. Climate and southern Africa's water–energy–food nexus. Nature Clim Change 5, 837–846. https://doi.org/10.1038/nclimate2735

Corbeels, M., Chirat, G., Messad, S., Thierfelder, C., 2016. Performance and sensitivity of the DSSAT crop growth model in simulating maize yield under conservation agriculture. European Journal of Agronomy 76, 41–53. https://doi.org/10.1016/j.eja.2016.02.001

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, 034011. https://doi.org/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 Biogeochemical Cycles 25. https://doi.org/10.1029/2009GB003765

Descheemaeker, K., Oosting, S.J., Tui, S.H.-K., Masikati, P., Falconnier, G.N., Giller, K.E., 2016. Climate change adaptation and mitigation in smallholder crop–livestock systems in sub-Saharan Africa: a call for integrated impact assessments. Reg Environ Change 16, 2331–2343. https://doi.org/10.1007/s10113-016-0957-8

Deutsch, C.A., Tewksbury, J.J., Tigchelaar, M., Battisti, D.S., Merrill, S.C., Huey, R.B., Naylor, R.L., 2018. Increase in crop losses to insect pests in a warming climate. Science 361, 916–919. https://doi.org/10.1126/science.aat3466

Donatelli, M., Magarey, R.D., Bregaglio, S., Willocquet, L., Whish, J.P.M., Savary, S., 2017. Modelling the impacts of pests and diseases on agricultural systems. Agric Syst 155, 213–224. https://doi.org/10.1016/j.agsy.2017.01.019

Durand, J.-L., Delusca, K., Boote, K., Lizaso, J., Manderscheid, R., Weigel, H.J., Ruane, A.C., Rosenzweig, C., Jones, J., Ahuja, L., Anapalli, S., Basso, B., Baron, C., Bertuzzi, P., Biernath, C., Deryng, D., Ewert, F., Gaiser, T., Gayler, S., Heinlein, F., Kersebaum, K.C., Kim, S.-H., Müller, C., Nendel, C., Olioso, A., Priesack, E., Villegas, J.R., Ripoche, D., Rötter, R.P., Seidel, S.I., Srivastava, A., Tao, F., Timlin, D., Twine, T., Wang, E., Webber, H., Zhao, Z., 2018. How accurately do maize crop models simulate the interactions of atmospheric CO2 concentration levels with limited water supply on water use and yield? European Journal of Agronomy,

Recent advances in crop modelling to support sustainable agricultural production and food security under global change 100, 67–75. https://doi.org/10.1016/j.eja.2017.01.002
Dzotsi, K.A., Jones, J.W., Adiku, S.G.K., Naab, J.B., Singh, U., Porter, C.H., Gijsman, A.J., 2010. Modeling soil and plant phosphorus within DSSAT. Ecological Modelling 221, 2839–2849. https://doi.org/10.1016/j.ecolmodel.2010.08.023
Falconnier, G.N., Descheemaeker, K., Van Mourik, T.A., Adam, M., Sogoba, B., Giller, K.E., 2017. Co-learning cycles to support the design of innovative farm systems in southern Mali. European Journal of Agronomy 89, 61–74. https://doi.org/10.1016/j.eja.2017.06.008
FAOSTAT, 2018. Food and Agriculture Organization of the United Nations Rome, Italy. http://faostat.fao.org.
Faye, B., Webber, H., Diop, M., Mbaye, M.L., Owusu-Sekyere, J.D., Naab, J.B., Gaiser, T., 2018a. Potential impact of climate change on peanut yield in Senegal, West Africa. Field Crops Research 219, 148–159. https://doi.org/10.1016/j.fcr.2018.01.034
Faye, B., Webber, H., Naab, J.B., MacCarthy, D.S., Adam, M., Ewert, F., Lamers, J.P.A., Schleussner, C.-F., Ruane, A., Gessner, U., Hoogenboom, G., Boote, K., Shelia, V., Saeed, F., Wisser, D., Hadir, S., Laux, P., Gaiser, T., 2018b. Impacts of 1.5 versus 2.0 °C on cereal yields in the West African Sudan Savanna. Environmental Research Letters 13, 034014. https://doi.org/10.1088/1748-9326/aaab40
Fleisher, D.H., Condori, B., Quiroz, R., Alva, A., Asseng, S., Barreda, C., Bindi, M., Boote, K.J., Ferrise, R., Franke, A.C., Govindakrishnan, P.M., Harahagazwe, D., Hoogenboom, G., Kumar, S.N., Merante, P., Nendel, C., Olesen, J.E., Parker, P.S., Raes, D., Raymundo, R., Ruane, A.C., Stockle, C., Supit, I., Vanuytrecht, E., Wolf, J., Woli, P., 2017. A potato model intercomparison across varying climates and productivity levels. Global Change Biology 23, 1258–1281. https://doi.org/10.1111/gcb.13411
Folberth, C., Yang, H., Gaiser, T., Abbaspour, K.C., Schulin, R., 2013. Modeling maize yield responses to improvement in nutrient, water and cultivar inputs in sub-Saharan Africa. Agricultural Systems 119, 22–34. https://doi.org/10.1016/j.agsy.2013.04.002
Franke, J., Müller, C., Elliott, J., Ruane, A.C., Jagermeyr, J., Balkovic, J., Ciais, P., Dury, M., Falloon, P., Folberth, C., Francois, L., Hank, T., Hoffmann, M., Izaurralde, R.C., Jacquemin, I., Jones, C., Khabarov, N., Koch, M., Li, M., Liu, W., Olin, S., Phillips, M., Pugh, T.A.M., Reddy, A., Wang, X., Williams, K., Zabel, F., Moyer, E., 2020. The GGCMI Phase II experiment: global gridded crop model simulations under uniform changes in CO_2, temperature, water, and nitrogen levels (protocol version 1.0). Geoscientific Model Development Discussions 1–30. https://doi.org/10.5194/gmd-2019-237

Freduah, B.S., MacCarthy, D.S., Adam, M., Ly, M., Ruane, A.C., Timpong-Jones, E.C., Traore, P.S., Boote, K.J., Porter, C., Adiku, S.G.K., 2019. Sensitivity of Maize Yield in Smallholder Systems to Climate Scenarios in Semi-Arid Regions of West Africa: Accounting for Variability in Farm Management Practices. Agronomy 9, 639. https://doi.org/10.3390/agronomy9100639

Frelat, R., Lopez-Ridaura, S., Giller, K.E., Herrero, M., Douxchamps, S., Djurfeldt, A.A., Erenstein, O., Henderson, B., Kassie, M., Paul, B.K., Rigolot, C., Ritzema, R.S., Rodriguez, D., van Asten, P.J.A., van Wijk, M.T., 2016. Drivers of household food availability in sub-Saharan Africa based on big data from small farms. Proceedings of the National Academy of Sciences of the United States of America 113, 458–463. https://doi.org/10.1073/pnas.1518384112

Gaiser, T., Perkons, U., Küpper, P.M., Kautz, T., Uteau-Puschmann, D., Ewert, F., Enders, A., Krauss, G., 2013. Modeling biopore effects on root growth and biomass production on soils with pronounced sub-soil clay accumulation. Ecological Modelling 256, 6–15. https://doi.org/10.1016/j.ecolmodel.2013.02.016

Guan, K., Sultan, B., Biasutti, M., Baron, C., Lobell, D.B., 2017. Assessing climate adaptation options and uncertainties for cereal systems in West Africa. Agricultural and Forest Meteorology 232, 291–305. https://doi.org/10.1016/j.agrformet.2016.07.021

Guntiñas, M.E., Leirós, M.C., Trasar-Cepeda, C., Gil-Sotres, F., 2012. Effects of moisture and temperature on net soil nitrogen mineralization: A laboratory study. European Journal of Soil Biology 48, 73–80. https://doi.org/10.1016/j.ejsobi.2011.07.015

Holzworth, D.P., Huth, N.I., deVoil, P.G., Zurcher, E.J., Herrmann, N.I., McLean, G., Chenu, K., van Oosterom, E.J., Snow, V., Murphy, C., Moore, A.D., Brown, H., Whish, J.P.M., Verrall, S., Fainges, J., Bell, L.W., Peake, A.S., Poulton, P.L., Hochman, Z., Thorburn, P.J., Gaydon, D.S., Dalgliesh, N.P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F.Y., Wang, E., Hammer, G.L., Robertson, M.J., Dimes, J.P., Whitbread, A.M., Hunt, J., van Rees, H., McClelland, T., Carberry, P.S., Hargreaves, J.N.G., MacLeod, N., McDonald, C., Harsdorf, J., Wedgwood, S., Keating, B.A., 2014. APSIM – Evolution towards a new generation of agricultural systems simulation. Environmental Modelling & Software 62, 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009

IPCC, 2013. Annex I: Atlas of Global and Regional Climate Projections, in: van Oldenborgh, G.J., Collins, J., Arblaster, J., Christensen, J., Marotzke, J., Power, S.B., Rummukainen, M., Zhou, T. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, US.

Jones, C.A., Kiniry, J.R., Dyke, P.T., 1986. CERES-Maize: a simulation model of maize growth and development. Texas AandM University Press.

Jones, James W., Naab, J., Fatondji, Dougbedji, Dzotsi, K., Adiku, S., He, J., 2012. Uncertainties in Simulating Crop Performance in Degraded Soils and Low Input Production Systems, in: Kihara, J.M., Fatondji, D., Jones, J. W., Hoogenboom, G., Tabo, R., Bationo, A. (Eds.), Improving Soil Fertility Recommendations in Africa Using the Decision Support System for Agrotechnology Transfer (DSSAT). Springer, Dordrecht, pp. 43–59. https://doi.org/10.1007/978-94-007-2960-5_4

Kamukondiwa, W., Bergström, L., 1994. Nitrate leaching in field lysimeters at an agricultural site in Zimbabwe. Soil Use and Management 10, 118–124. https://doi.org/10.1111/j.1475-2743.1994.tb00471.x

Kersebaum, K.C., 2011. Special Features of the HERMES Model and Additional Procedures for Parameterization, Calibration, Validation, and Applications, in: Methods of Introducing System Models into Agricultural Research. L.R. Ahuja and L. Ma, pp. 65–94.

Kersebaum, K.C., Boote, K.J., Jorgenson, J.S., Nendel, C., Bindi, M., Frühauf, C., Gaiser, T., Hoogenboom, G., Kollas, C., Olesen, J.E., Rötter, R.P., Ruget, F., Thorburn, P.J., Trnka, M., Wegehenkel, M., 2015. Analysis and classification of data sets for calibration and validation of agro-ecosystem models. Environmental Modelling & Software 72, 402–417. https://doi.org/10.1016/j.envsoft.2015.05.009

Kihara, J., Fatondji, D., Jones, J.W., Hoogenboom, G., Tabo, R., Bationo, A. (Eds.), 2012. Improving Soil Fertility Recommendations in Africa using the Decision Support System for Agrotechnology Transfer (DSSAT). Springer Netherlands. https://doi.org/10.1007/978-94-007-2960-5

Kim, S.-H., Yang, Y., Timlin, D.J., Fleisher, D.H., Dathe, A., Reddy, V.R., Staver, K., 2012. Modeling Temperature Responses of Leaf Growth, Development, and Biomass in Maize with MAIZSIM. Agronomy Journal 104, 1523. https://doi.org/10.2134/agronj2011.0321

Knutti, R., 2010. The end of model democracy? Climatic Change 102, 395–404. https://doi.org/10.1007/s10584-010-9800-2

Leakey, A.D.B., Uribelarrea, M., Ainsworth, E.A., Naidu, S.L., Rogers, A., Ort, D.R., Long, S.P., 2006. Photosynthesis, Productivity, and Yield of Maize Are Not Affected by Open-Air Elevation of CO2 Concentration in the Absence of Drought. Plant Physiology 140, 779–790. https://doi.org/10.1104/pp.105.073957

Li, T., Hasegawa, T., Yin, X., Zhu, Y., Boote, K., Adam, M., Bregaglio, S., Buis, S., Confalonieri, R., Fumoto, T., Gaydon, D., Marcaida, M., Nakagawa, H., Oriol, P., Ruane, A.C., Ruget, F., Singh, B.-, Singh, U., Tang, L., Tao, F., Wilkens, P., Yoshida, H., Zhang, Z., Bouman, B., 2015. Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. Glob Chang Biol 21, 1328–1341. https://doi.org/10.1111/gcb.12758

Lizaso, J.I., Boote, K.J., Jones, J.W., Porter, C.H., Echarte, L., Westgate, M.E., Sonohat, G., 2011. CSM-IXIM: A New Maize Simulation Model for DSSAT Version 4.5. Agronomy Journal 103, 766–779. https://doi.org/10.2134/agronj2010.0423

Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L., 2008. Prioritizing Climate Change Adaptation Needs for Food Security in 2030. Science 319, 607–610. https://doi.org/10.1126/science.1152339

Loon, M.P. van, Hijbeek, R., Berge, H.F.M. ten, Sy, V.D., Broeke, G.A. ten, Solomon, D., Ittersum, M.K. van, 2019. Impacts of intensifying or expanding cereal cropping in sub-Saharan Africa on greenhouse gas emissions and food security. Global Change Biology 25, 3720–3730. https://doi.org/10.1111/gcb.14783

MacCarthy, D.S., Akponikpe, P.B.I., Narh, S., Tegbe, R., 2015. Modeling the effect of seasonal climate variability on the efficiency of mineral fertilization on maize in the coastal savannah of Ghana. Nutr Cycl Agroecosyst 102, 45–64. https://doi.org/10.1007/s10705-015-9701-x

Manderscheid, R., Erbs, M., Weigel, H.-J., 2014. Interactive effects of free-air CO2 enrichment and drought stress on maize growth. European Journal of Agronomy, Land, Climate and Resources 2020. Decision Support for Agriculture under Climate Change 52, 11–21. https://doi.org/10.1016/j.eja.2011.12.007

Mapanda, F., Wuta, M., Nyamangara, J., Rees, R.M., 2012. Nitrogen leaching and indirect nitrous oxide emissions from fertilized croplands in Zimbabwe. Nutr Cycl Agroecosyst 94, 85–96. https://doi.org/10.1007/s10705-012-9528-7

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J.W., Rötter, R.P., Boote, K.J., Ruane, A.C., Thorburn, P.J., Cammarano, D., Hatfield, J.L., Rosenzweig, C., Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R.F., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Kumar, S.N., Nendel, C., O'leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2015. Multimodel ensembles of wheat growth: many models are better than one. Global Change Biology 21, 911–925. https://doi.org/10.1111/gcb.12768

Masvaya, E.N., Nyamangara, J., Descheemaeker, K., Giller, K.E., 2017. Tillage, mulch and fertiliser impacts on soil nitrogen availability and maize production in semi-arid Zimbabwe. Soil and Tillage Research 168, 125–132. https://doi.org/10.1016/j.still.2016.12.007

Ndoli, A., Baudron, F., Sida, T.S., Schut, A.G.T., van Heerwaarden, J., Giller, K.E., 2018. Conservation agriculture with trees amplifies negative effects of reduced tillage on maize performance in East Africa. Field Crops Research 221, 238–244. https://doi.org/10.1016/j.fcr.2018.03.003

Nendel, C., Berg, M., Kersebaum, K.C., Mirschel, W., Specka, X., Wegehenkel, M., Wenkel, K.O., Wieland, R., 2011. The MONICA model: Testing predictability for crop growth, soil moisture and nitrogen dynamics. Ecological Modelling 222, 1614–1625. https://doi.org/10.1016/j.ecolmodel.2011.02.018

Nendel, C., Melzer, D., Thorburn, P.J., 2019. The nitrogen nutrition potential of arable soils. Sci Rep 9, 1–9. https://doi.org/10.1038/s41598-019-42274-y

OCDE, FAO, 2016. Agriculture in Sub-Saharan Africa: Prospects and challenges for the next decade, in: OECD-FAO Agricultural Outlook 2016-2025. Paris.

Ricome, A., Affholder, F., Gérard, F., Muller, B., Poeydebat, C., Quirion, P., Sall, M., 2017. Are subsidies to weather-index insurance the best use of public funds? A bio-economic farm model applied to the Senegalese groundnut basin. Agricultural Systems 156, 149–176. https://doi.org/10.1016/j.agsy.2017.05.015

Ritchie, J.T., Singh, U., Godwin, D.C., Bowen, W.T., 1998. Cereal growth, development and yield, in: Tsuji, G.Y., Hoogenboom, G., Thornton, P.K. (Eds.), Understanding Options for Agricultural Production, Systems Approaches for Sustainable Agricultural Development. Springer Netherlands, Dordrecht, pp. 79–98. https://doi.org/10.1007/978-94-017-3624-4_5

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. PNAS 111, 3268–3273. https://doi.org/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. Agricultural and Forest Meteorology, Agricultural prediction using climate model ensembles 170, 166–182. https://doi.org/10.1016/j.agrformet.2012.09.011

Rötter, R., van Keulen, H., Jansen, M.J.W., 1997. Variations in yield response to fertilizer application in the tropics: I. Quantifying risks and opportunities for smallholders based on crop growth simulation. Agricultural Systems 53, 41–68. https://doi.org/10.1016/S0308-521X(96)00036-4

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

Ruane, A.C., Rosenzweig, C., Asseng, S., Boote, K.J., Elliott, J., Ewert, F., Jones, J.W., Martre, P., McDermid, S.P., Müller, C., Snyder, A., Thorburn, P.J., 2017. An AgMIP framework for improved agricultural representation in integrated assessment models. Environ. Res. Lett. 12, 125003. https://doi.org/10.1088/1748-9326/aa8da6

Rurinda, J., van Wijk, M.T., Mapfumo, P., Descheemaeker, K., Supit, I., Giller, K.E., 2015. Climate change and maize yield in southern Africa: what can farm management do? Glob Change Biol 21, 4588–4601. https://doi.org/10.1111/gcb.13061

Russo, T.A., Tully, K., Palm, C., Neill, C., 2017. Leaching losses from Kenyan maize cropland receiving different rates of nitrogen fertilizer. Nutr Cycl Agroecosyst 108, 195–209. https://doi.org/10.1007/s10705-017-9852-z

Sadhukhan, D., Qi, Z., Zhang, T., Tan, C.S., Ma, L., Andales, A.A., 2019. Development and evaluation of a phosphorus (P) module in RZWQM2 for phosphorus management in agricultural fields. Environmental Modelling & Software 113, 48–58. https://doi.org/10.1016/j.envsoft.2018.12.007

Sida, T.S., Baudron, F., Hadgu, K., Derero, A., Giller, K.E., 2018. Crop vs. tree: Can agronomic management reduce trade-offs in tree-crop interactions? Agriculture, Ecosystems & Environment 260, 36–46. https://doi.org/10.1016/j.agee.2018.03.011

Smith, W., Grant, B., Qi, Z., He, W., VanderZaag, A., Drury, C.F., Helmers, M., 2020. Development of the DNDC model to improve soil hydrology and incorporate mechanistic tile drainage: A comparative analysis with RZWQM2. Environmental Modelling & Software 123, 104577. https://doi.org/10.1016/j.envsoft.2019.104577

Stitt, M., Krapp, A., 1999. The interaction between elevated carbon dioxide and nitrogen nutrition: the physiological and molecular background. Plant, Cell & Environment 22, 583–621. https://doi.org/10.1046/j.1365-3040.1999.00386.x

Sultan, B., Gaetani, M., 2016. Agriculture in West Africa in the Twenty-First Century: Climate Change and Impacts Scenarios, and Potential for Adaptation. Front. Plant Sci. 7. https://doi.org/10.3389/fpls.2016.01262

Sultan, B., Guan, K., Kouressy, M., Biasutti, M., Piani, C., Hammer, G.L., McLean, G., Lobell, D.B., 2014. Robust features of future climate change impacts on sorghum yields in West Africa. Environ. Res. Lett. 9, 104006. https://doi.org/10.1088/1748-9326/9/10/104006

Tao, F., Palosuo, T., Rötter, R.P., Díaz-Ambrona, C.G.H., Inés Mínguez, M., Semenov, M.A., Kersebaum, K.C., Cammarano, D., Specka, X., Nendel, C., Srivastava, A.K., Ewert, F., Padovan, G., Ferrise, R., Martre, P., Rodríguez, L., Ruiz-Ramos, M., Gaiser, T., Höhn, J.G., Salo, T., Dibari, C., Schulman, A.H., 2020. Why do crop models diverge substantially in climate impact projections? A comprehensive analysis based on eight barley crop models. Agricultural and Forest Meteorology 281, 107851. https://doi.org/10.1016/j.agrformet.2019.107851

Tao, F., Rötter, R.P., Palosuo, T., Díaz-Ambrona, C.G.H., Mínguez, M.I., Semenov, M.A., Kersebaum, K.C., Nendel, C., Specka, X., Hoffmann, H., Ewert, F., Dambreville, A., Martre, P., Rodríguez, L., Ruiz-Ramos, M., Gaiser, T., Höhn, J.G., Salo, T., Ferrise, R., Bindi, M., Cammarano, D., Schulman, A.H., 2018. Contribution of crop model structure, parameters and climate projections to uncertainty in climate change impact assessments. Global Change Biology 24, 1291–1307. https://doi.org/10.1111/gcb.14019

Tao, F., Zhang, Z., 2010. Adaptation of maize production to climate change in North China Plain: Quantify the relative contributions of adaptation options. European Journal of Agronomy 33, 103–116. https://doi.org/10.1016/j.eja.2010.04.002

Taylor, C.M., Belušić, D., Guichard, F., Parker, D.J., Vischel, T., Bock, O., Harris, P.P., Janicot, S., Klein, C., Panthou, G., 2017. Frequency of extreme Sahelian storms tripled since 1982 in satellite observations. Nature 544, 475–478. https://doi.org/10.1038/nature22069

Taylor, S.L., Payton, M.E., Raun, W.R., 1999. Relationship between mean yield, coefficient of variation, mean square error, and plot size in wheat field experiments. Communications in Soil Science and Plant Analysis 30, 1439–1447. https://doi.org/10.1080/00103629909370298

ten Berge, H.F.M., Hijbeek, R., van Loon, M.P., Rurinda, J., Tesfaye, K., Zingore, S., Craufurd, P., van Heerwaarden, J., Brentrup, F., Schröder, J.J., Boogaard, H.L., de Groot, H.L.E., van Ittersum, M.K., 2019. Maize crop nutrient input requirements for food security in sub-Saharan Africa. Global Food Security 23, 9–21. https://doi.org/10.1016/j.gfs.2019.02.001

Traore, B., Descheemaeker, K., van Wijk, M.T., Corbeels, M., Supit, I., Giller, K.E., 2017. Modelling cereal crops to assess future climate risk for family food self-sufficiency in southern Mali. Field Crops Research 201, 133–145. https://doi.org/10.1016/j.fcr.2016.11.002

Traore, B., van Wijk, M.T., Descheemaeker, K., Corbeels, M., Rufino, M.C., Giller, K.E., 2014. Evaluation of climate adaptation options for Sudano-Sahelian cropping systems. Field Crops Research 156, 63–75. https://doi.org/10.1016/j.fcr.2013.10.014

Twine, T.E., Bryant, J.J., Richter, K.T., Bernacchi, C.J., McConnaughay, K.D., Morris, S.J., Leakey, A.D.B., 2013. Impacts of elevated CO2 concentration on the productivity and surface energy budget of the soybean and maize agroecosystem in the Midwest USA. Global Change Biology 19, 2838–2852. https://doi.org/10.1111/gcb.12270

van der Laan, M., Stirzaker, R.J., Annandale, J.G., Bristow, K.L., Preez, C.C. du, 2010. Monitoring and modelling draining and resident soil water nitrate concentrations to estimate leaching losses. Agricultural Water Management 97, 1779–1786. https://doi.org/10.1016/j.agwat.2010.06.012

Vanlauwe, B., Coyne, D., Gockowski, J., Hauser, S., Huising, J., Masso, C., Nziguheba, G., Schut, M., Van Asten, P., 2014. Sustainable intensification and the African smallholder farmer. Current Opinion in Environmental Sustainability 8, 15–22. https://doi.org/10.1016/j.cosust.2014.06.001

Velde, M. van der, Folberth, C., Balkovič, J., Ciais, P., Fritz, S., Janssens, I.A., Obersteiner, M., See, L., Skalský, R., Xiong, W., Peñuelas, J., 2014. African crop yield reductions due to increasingly unbalanced Nitrogen and Phosphorus consumption. Global Change Biology 20, 1278–1288. https://doi.org/10.1111/gcb.12481

Waha, K., Müller, C., Rolinski, S., 2013. Separate and combined effects of temperature and precipitation change on maize yields in sub-Saharan Africa for mid- to late-21st century. Global and Planetary Change 106, 1–12. https://doi.org/10.1016/j.gloplacha.2013.02.009

Wallach, D., Martre, P., Liu, B., Asseng, S., Ewert, F., Thorburn, P.J., Ittersum, M. van, Aggarwal, P.K., Ahmed, M., Basso, B., Biernath, C., Cammarano, D., Challinor, A.J., Sanctis, G.D., Dumont, B., Rezaei, E.E., Fereres, E., Fitzgerald, G.J., Gao, Y., Garcia-Vila, M., Gayler, S., Girousse, C., Hoogenboom, G., Horan, H., Izaurralde, R.C., Jones, C.D., Kassie, B.T., Kersebaum, K.C., Klein, C., Koehler, A.-K., Maiorano, A., Minoli, S., Müller, C., Kumar, S.N., Nendel, C., O'Leary, G.J., Palosuo, T., Priesack, E., Ripoche, D., Rötter, R.P., Semenov, M.A., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Wolf, J., Zhang, Z., 2018. Multimodel ensembles improve predictions of crop–environment–management interactions. Global Change Biology 24, 5072–5083. https://doi.org/10.1111/gcb.14411

Wood, S.A., Tirfessa, D., Baudron, F., 2018. Soil organic matter underlies crop nutritional quality and productivity in smallholder agriculture. Agriculture, Ecosystems & Environment 266, 100–108. https://doi.org/10.1016/j.agee.2018.07.025

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D.B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J.-L., Elliott, J., Ewert, F., Janssens, I.A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C., Peng, S., Peñuelas, J., Ruane, A.C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., Asseng, S., 2017. Temperature increase reduces global yields of major crops in four independent estimates. PNAS 114, 9326–9331. https://doi.org/10.1073/pnas.1701762114

Ziska, L.H., Sicher, R.C., Bunce, J.A., 1999. The impact of elevated carbon dioxide on the growth and gas exchange of three C4 species differing in CO2 leak rates. Physiologia Plantarum 105, 74–80. https://doi.org/10.1034/j.1399-3054.1999.105112.x

Ziska, L.H., Weerakoon, W., Namuco, O.S., Pamplona, R., 1996. The Influence of Nitrogen on the Elevated CO2 Response in Field-Grown Rice. Functional Plant Biol. 23, 45–52. https://doi.org/10.1071/pp9960045

Figures:

Figure 1: (A), Map of sub-Saharan Africa showing the five study sites representative of FAO tropical agro-ecological zones where maize cultivation is possible. (B), Observed soil water content to maximum rooting depth in the six experiments where soil water was monitored (vertical lines from left to right are sowing, anthesis and maturity dates). PAWC: Plant Available Soil Water Capacity. (C), Observed maize grain yield at the five sites for two growing seasons (ten experiments) as a function of estimated available mineral nitrogen (N), *i.e.* the summation of initial soil mineral N, applied mineral and organic N and mineralized soil organic N and manure N over the whole growing season (for Ethiopia and Rwanda, initial mineral N measurements were not available

Figure 2: (A), Observed (crosses with standard deviation if known) and simulated (box plots) grain yields. Simulations are from an ensemble of 25 partially and fully calibrated models. The line in the box and the width of the box are the median and the interquartile range respectively. The whiskers extend from the edge of the box to the most extreme data point below 1.5 interquartile range. Black open dots are outliers. (B) rRMSE (averaged across all models) of simulated – observed comparison for six variables of interest. For aboveground plant nitrogen the comparison was possible for four of the ten experiments only (Benin and Ghana). Open dots indicate rRMSE of ensemble median. (C) Coefficient of variation (averaged across sites) of 25 model simulations for five variables.

Figure 3: Relative variation (mean ±standard deviation) between average of n models and measured grain yield in the ten experiments at five sites across sub-Saharan Africa. Models were randomly selected among the 25 calibrated models that simulated yield for the ten experiments. The horizontal dotted line is the 13.5% threshold, *i.e.* the coefficient of variation for measured yields typically obtained in experimental plots (Taylor et al., 1999).

Figure 4: Boxplots of relative change in grain yield (compared with baseline climate) when doubling [CO_2], increasing temperature by +4°C, increasing and decreasing rainfall (150% and 50% of baseline) in five sites across sub-Saharan Africa and for three N inputs of 0, 80 and 160 kg N ha^{-1}. Simulations are from 24 maize models with full calibration (one model did not perform the sensitivity analysis). Two models not simulating the effect of N on crop growth are displayed only for 160 kg N ha^{-1}. The line in the box and the width of the box are the median and the interquartile range respectively. The whiskers extend from the edge of the box to the most extreme data point below 1.5 interquartile range. Outliers (data points below Q1 – 1.5×(Q3-Q1) or above Q3 + 1.5×(Q3-Q1) where Q1 is the first quartile and Q3 the third quartile) were not displayed.

Figure 5: Mean (± SE) relative change in grain yield (compared with baseline climate) when doubling [CO_2], increasing temperature by +4°C, increasing and decreasing rainfall (150% and 50% of current) in five sites across sub-Saharan Africa and for three N inputs of 0, 80 and 160 kg N ha^{-1}. Simulations are from 25 maize models with full calibration classified in three classes: two models that did not simulate responses to N inputs (class 1, red), three models that simulated response to N inputs but without a daily N module (Class 2, green) and with a daily N module (Class 3). Models below the median sum of ranks for rRMSE over all the simulated variables were classified as "most consistent" models (class 3a, cyan), models above the median as "less consistent" models (class 3b, purple) (see section 2.3 for a detailed description of the classification). The reader is referred to the web version of this article for interpretation of references to colors.

Tables

Table 1: Characteristics of the five sites and ten experiments selected for model evaluation of maize yield simulation in rainfed smallholder farming systems across sub-Saharan Africa.

		Site				
General information	Country	**Benin**	**Mali**	**Ghana**	**Rwanda**	**Ethiopia**
	Location	Ouri Yori	Ntarla	Kpong	Bugesera	Bako
	Source	(Amouzou et al., 2018)	(Traore et al., 2014)	(MacCarthy et al., 2015)	(Ndoli et al., 2018)	(Sida et al., 2018)
	Latitude	10.82	12.58	6.16	-2.35	9.13
	Longitude	1.07	-5.70	0.06	30.27	37.10
	Elevation (m)	213	302	22	1400	1700-2000
	FAO AEZ	Tropic - warm semi-arid	Tropic - warm semi-arid	Tropic - warm sub-humid	Tropic - cool sub-humid	Tropic - cool - subhumid
	Rainfall pattern	unimodal	unimodal	bimodal	bimodal	unimodal
	Average growing season	june-september	june-september	march-july and august-december[1]	september-january and february-july[2]	june-october
Soils	Soil type (FAO)	Gleyic Alisol	Ferric Lixisol	Vertisol	Humic Ferralsol	Nitisol
	Soil texture	loamy sand	loamy sand	clay	sandy loam	clay
	maximum rooting depth (cm)	60	120	100	100	120
	Plant Available Soil Water Capacity (mm to maximum rooting depth)	105	167	93	104	202
	SOC (%) (0-30cm)	0.28	0.2	0.57	1.65	0.65
	Total Nitrogen (%) (0-30cm)	0.023	0.015	0.048	0.138	0.059
Management	Cultivar	EVDT-97 STR (OPV[3])	Suwan 1 - SR (OPV[3])	Obatampa (OPV[3])	ZM607 (OPV[3])	BH540 (Hybrid)
	Sowing dates (DOY[4])	176, 185	151, 163	111, 105	282, 267	161, 158
	Manure input (t/ha)	0	3	0	0	0
	N content in manure (%)	-	1.6	-	-	-
	Total applied N fertiliser (kgN/ha)	0	85	80	64	87
	Total applied P fertiliser (kgP/ha)	0	26	30	20	9
	Total applied K fertilister (kgK/ha)	0	16	37	0	0
Phenology	Anthesis (DAP[5])	52, 53	56, 54	65, 60	72, 82	97, 98
	Maturity (DAP[5])	80, 86	97, 95	105, 106	120, 118	138, 139
Experimental year climate	Experimental year (first experiment)	2014	2009	2008	2013-2014	2013
	Mean growing season	27.9	26.6	27.7	22.8	21.1

	temperature					
	Mean growing season precipitation	516	549	536	217	476
	Experimental year (second experiment)	2015	2010	2009	2014-2015	2014
	Mean growing season temperature (season 2)	27.1	26.9	27.6	23.1	20.5
	Mean growing season precipitation (season 2)	810	705	455	351	923
Baseline climate (1980-2010)	Mean growing season temperature	25.5	28.3	27.6	21.9	20.6
	Mean growing season precipitation	641	582	442	331	939

[1]Only March-July was considered for the experiments

[2]Only September-January was considered for the experiments

[3]Open pollinated variety

[4]Day of the year. First and second value indicate season 1 and season 2 experiments, respectively.

[5]Days after planting. First and second value indicate season 1 and season 2 experiments, respectively.

Table 2: Model grouping into four groups according to characteristics linked to the simulation of N and additional characteristics of the models. Class 3a and 3b were determined after the analysis of model ranking (based on rRMSE) when simulating all variables of interest (see section 2.4 for detailed description of the classification). In bold, models that participated in a previous maize intercomparison in high input systems (Bassu et al., 2014). Underlined models are the ten highest ranked models (among class 3 models) for grain and biomass simulation (see section 2.4 for detailed description of the classification).

Model Class	effect of N input	Daily N module	Model	Model reference*	Leaf area development and light interception[a]	Light utilization[b]	Yield formation[c]	Crop phenology[d]	Root distribution over depth[e]	Simulation of N leaching	Simulation of heat stress	Type of water stress[f]	Type of heat stress[g]	Water dynamics[h]	Evapotranspiration[i]	Soil CN model[j]	Process modified by elevated CO2[k]
1	no	no	GLAM	Challinor et al. (2004)	S	RUE,TE	B, HI	T, DL	LIN	no	yes	E	R	C	PT	-	RUE, TE
			MCWLA	Tao and Zhang (2010)	S	P-R	B, HI	T	EXP	no	yes	E	V,R	R	PM	-	-
2	yes	no	**PEGASUS**	Deryng et al. (2014)	S	RUE	B, Prt	T	LIN	no	yes	E, S	V,R	C	PT	C, P(1)	RUE, TE
			SARRA-H	Baron et al. (2005)	S	RUE	HI, Prt	T	LIN	no	no	S	-	C	PM	-	RUE, T
			CELSIUS	Ricome et al. (2017)	S	RUE	B, Gn, Hi_mw	T, DL	LIN	no	yes	S	V,R	C	PM	N	RUE
3a	yes	yes	**APSIM 7.9**	Holzworth et al. (2014)	S	RUE	Prt	T, DL	EXP	yes	yes	S	V	C	PT	CN, P(3), B	RUE, TE
			DNDC	Smith et al. (2020)	S	TE	HI	T	EXP	yes	yes	S	R	C	PM	CN,P(5),B	PT
			HERMES	Kersebaum (2011)	D	P-R	Prt	T, DL, O	EXP	yes	no	E, S	-	C	PM	N, P(2)	LF, T
			DSSAT-IXIM-Maize+Century	Lizaso et al. (2011)	D	P-R	Gn	T, DL	EXP	yes	yes	E	R	C	PT	CN, P(2), B	RUE, T
			DSSAT-IXIM-Maize+Ceres-SOM	Lizaso et al. (2011)	D	P-R	Gn	T, DL	EXP	yes	yes	E	R	C	PT	CN, P(1)	RUE, T
			MONICA	Nendel et al. (2011)	D	P-R	Prt	T, DL, O	EXP	yes	yes	E	V	C	PM	CN, P(6), B	-
			SALUS	Basso et al. (2010)	S	RUE	HI, Prt	T, DL	EXP	yes	yes	E	V	C	PT	CN, P(3), B	-
			SIMPLACE-Lintul + ET Hargreaves + Heat stress with air temperature	Gaiser et al. (2013)	S	RUE	Prt	T, DL	EXP	yes	no	E, S	-	C	O	CN, P(7), B	RUE, TE
			STICS	Brisson et al. (2002)	S	RUE	B, Gn, HI,mw	T, DL, O	SIG	yes	yes	E	V,R	C	SW	CN, P(2), B	RUE, T
			DSSAT-CERES-Maize+Century	Ritchie et al. (1998)	S	RUE	Gn	T, DL	EXP	yes	yes	E	R	C	PT	CN, P(2), B	RUE, T
3b	yes	yes	AGRO-IBIS	Twine et al. (2013)	S	P-R	B, Prt	T	EXP	yes	yes	S	V,R	R	O	C, N, P(2)	F
			APSIM 7.10	Holzworth et al. (2014)	S	RUE	Prt	T, DL	EXP	yes	yes	S	V	C	PT	CN, P(3), B	RUE, TE

DSSAT-CERES-Maize+Ceres-SOM	Ritchie et al. (1998)	S	RUE	Gn	T, DL	EXP	yes	yes	E	R	C	PT	CN, P(1)	RUE, T
EXPERT-N-Ceres	Biernath et al. (2011)	S	RUE	B, Gn	T, DL	EXP	yes	yes	E, S	V	R	PM	CN, P(3), B	-
EXPERT-N-Spass	Biernath et al. (2011)	D	P-R	Prt	T, DL	EXP	yes	yes	E, S	V	R	PM	CN, P(3), B	-
EXPERT-N-Sucros	Biernath et al. (2011)	D	P-R	Prt	T	EXP	yes	yes	E, S	V	R	PM	CN, P(3), B	-
MAIZSIM	Kim et al. (2012)	D	P-R	HI, Prt	T, DL	CD	yes	yes	O	V,R	R	P, O	N, P(1), B	LF, T, F
RZWQM2	Sadhukhan et al. (2019)	S	RUE	B, Gn, Prt	T, DL, O	EXP	yes	yes	E, S	V,R	R	SW	C, N, P(1), B	PT
SIMPLACE-Lintul + ET FAO-56 + Heat stress with crop temperature	Faye et al. (2018a)	S	RUE	Prt	T, DL	EXP	yes	yes	E, S	R	C	PM	CN, P(7), B	RUE, TE
SWB	van der Laan et al. (2010)	S	RUE,TE	Prt	T	LIN	yes	no	S	-	C	PM	CN, P(4)	RUE, TE

[a] *S, Simple-unilayer(e.g. LAI); D, Detailed Multilayer (e.g. canopy layers)*

[b] *RUE, radiation use efficiency approach; P-R gross photosynthesis - respiration; TE, compute water use first, then biomass growth from transpiration efficiency*

[c] *HI, fixed harvest index; B, total (above-ground) biomass; Gn, number of grains; Prt, partitioning during reproductive stage; HI_mw, Harvest Index modified by water stress*

[d] *Function of : T, Temperature; DL, photoperiod (day length); O, other water/nutrient stress effects considered*

[e] *LIN, Linear; EXP, Exponential; SIG, sigmoidal ;, CD, Convective Dispersive*

[f] *E = Eta/Etp, S = soil available water in root zone, O, leaf energy balance, leaf and soil water potential effects on photosynthesis and leaf expansion*

[g] *V = vegetative (source), R = reproductive organ (sink).*

[h] *C, 'Tipping bucket' capacity approach; R, Richards approach*

[i] *P, Penman; PM, Penman-Monteith; PT, Priestley –Taylor; SW, Shuttleworth-Wallace, O, leaf energy balance (MZ), Hargreaves Dual crop coefficient method (SI2), water demand in plant, root water uptake, closes surface energy budget (AG).*

[j] *C, C model; N, N model; P(x), x number of organic matter pools; B, microbial biomass pool.*

[k] *LF, Leaf-level photosynthesis-rubisco or on QE and Amax; RUE, Radiation use efficiency; TE, Transpiration efficiency; PT, Photosynthesis and transpiration ;F, Farquhar model, GY, Grain Yield; T, Stomatal conductance.*

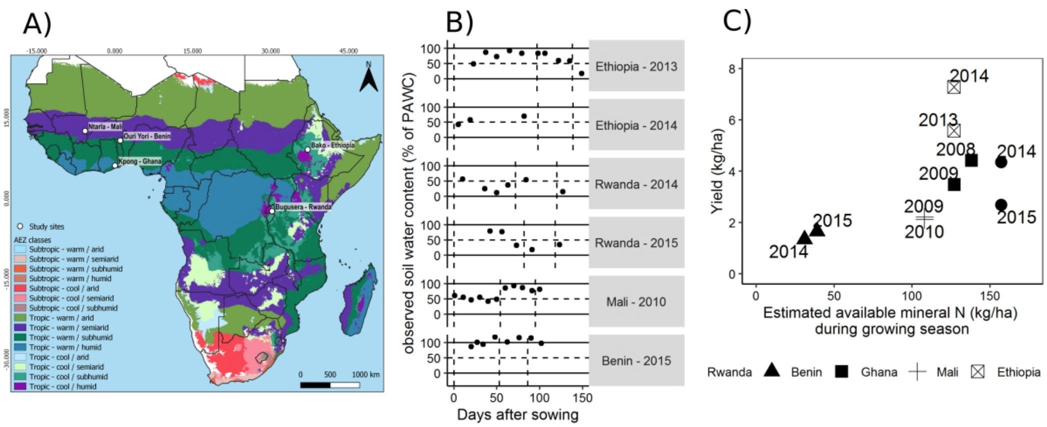
*More references and model documentation can be found in Table S2.

Table 3: rRMSE of simulated – observed comparison for six variables of interest for 25 fully calibrated maize models. In bold, models below median sum of ranks for all variables or yield and biomass only. Five models without daily simulation of N dynamics were not ranked.

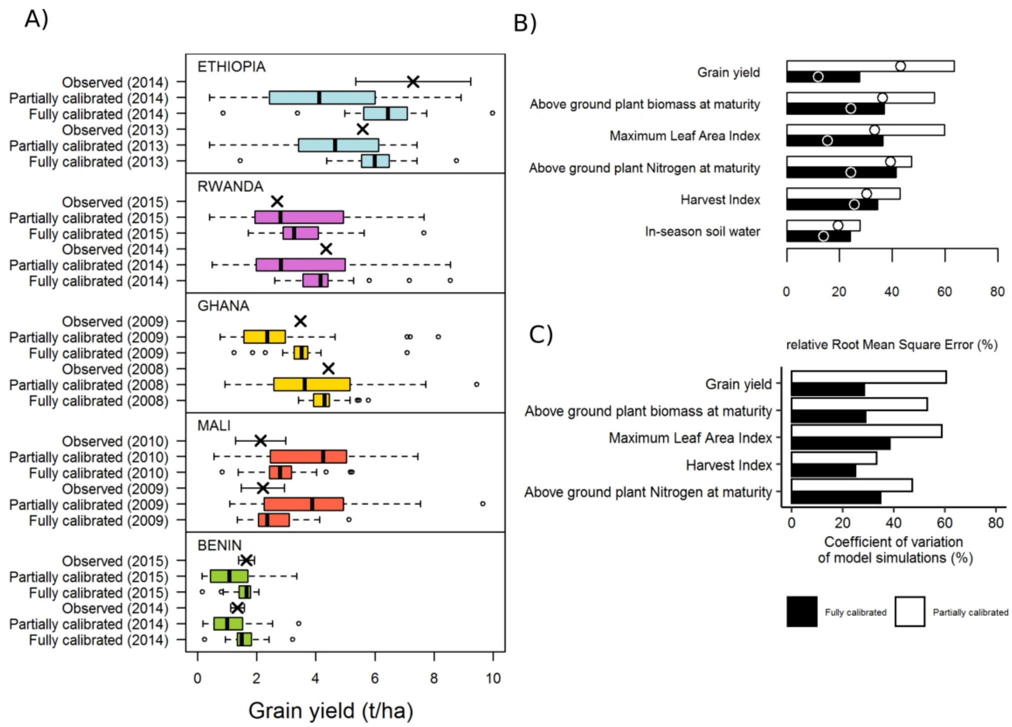
		rRMSE (%)						Rank			
Model class	Model name	grain yield	total above ground biomass	maximum LAI	total above ground plant N	Harvest Index	Soil water	Sum of ranks (all variables)	Sum of ranks (yield and biomass)	Final rank (all variables)	Final rank (yield and biomass)
1	GLAMM	18	31	57	-	32	-	-	-	-	-
	MCWLA	8	41	15	-	32	13	-	-	-	-
2	CELSIUS	12	26	34	-	33	12	-	-	-	-
	SARRA-H	17	31	10	-	34	17	-	-	-	-
	PEGASUS	16	43	79	-	57	78	-	-	-	-
3a	DNDC	22	34	40	7	21	9	32	20	**1**	**9**
	STICS	8	26	13	52	23	22	42	6	**2**	**1**
	HERMES	23	17	48	26	27	12	43	11	**3**	**4**
	DSSAT-IXIM-Maize+Century	20	25	46	28	29	17	47	10	**4**	**3**
	APSIM v 7.9	27	27	40	30	31	14	48	18	**5**	**7**
	DSSAT-IXIM-Maize+Ceres-SOM	21	28	41	33	29	17	51	15	**6**	**5**
	SIMPLACE-Lintul + Option 2*	11	30	6	43	38	24	54	10	**7**	**2**
	MONICA	42	46	11	15	30	18	60	35	**8**	16
	SALUS	36	48	6	11	41	23	73	35	**9**	15
	DSSAT-CERES-Maize+Century	34	33	58	52	31	14	75	24	**10**	**10**
3b	MAIZSIM	40	32	41	44	32	22	77	27	11	12
	APSIM v7.10	15	36	41	58	33	26	78	17	12	**6**
	DSSAT-CERES-Maize+Ceres-SOM	36	35	58	56	32	14	84	27	13	11
	SIMPLACE-Lintul + Option 1**	42	48	45	30	34	18	85	38	14	18
	EXPERT-N-Sucros	19	36	47	78	39	29	93	20	15	**8**
	RZWQM2	40	54	48	50	41	19	99	37	16	17
	SWB	72	58	37	49	37	43	99	42	17	20
	EXPERT-N-Spass	29	46	43	56	54	32	103	28	18	13
	AGRO-IBIS	82	47	28	51	50	72	104	39	19	19
	EXPERT-N-Ceres	27	67	61	96	41	30	116	33	20	14

*ET Hargreaves + Heat stress with air temperature

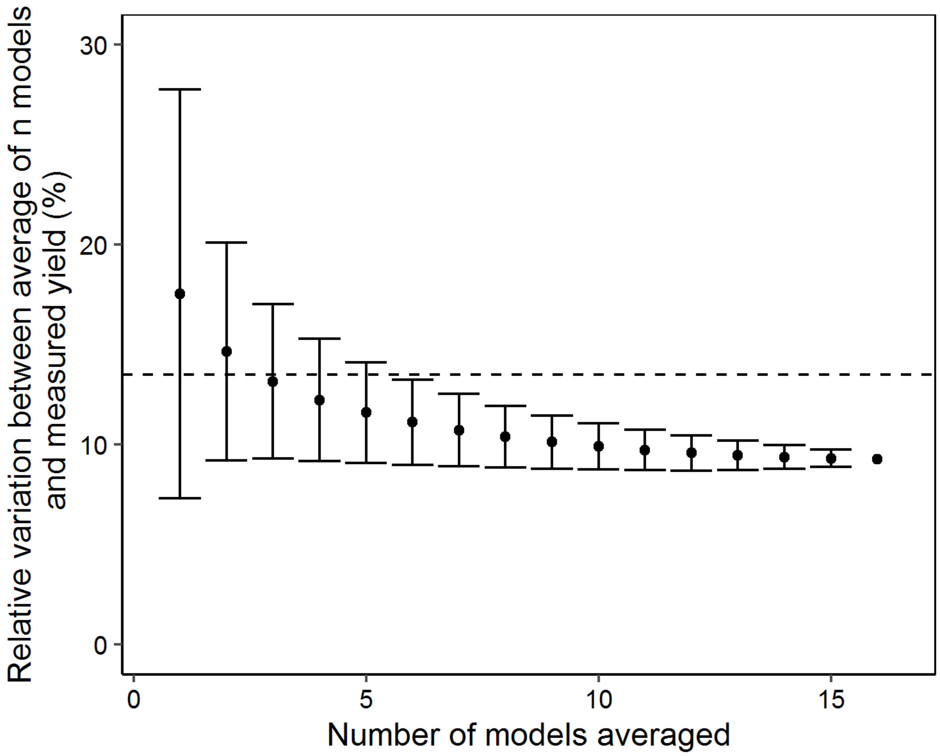
**ET FAO-56 + Heat stress with crop temperature



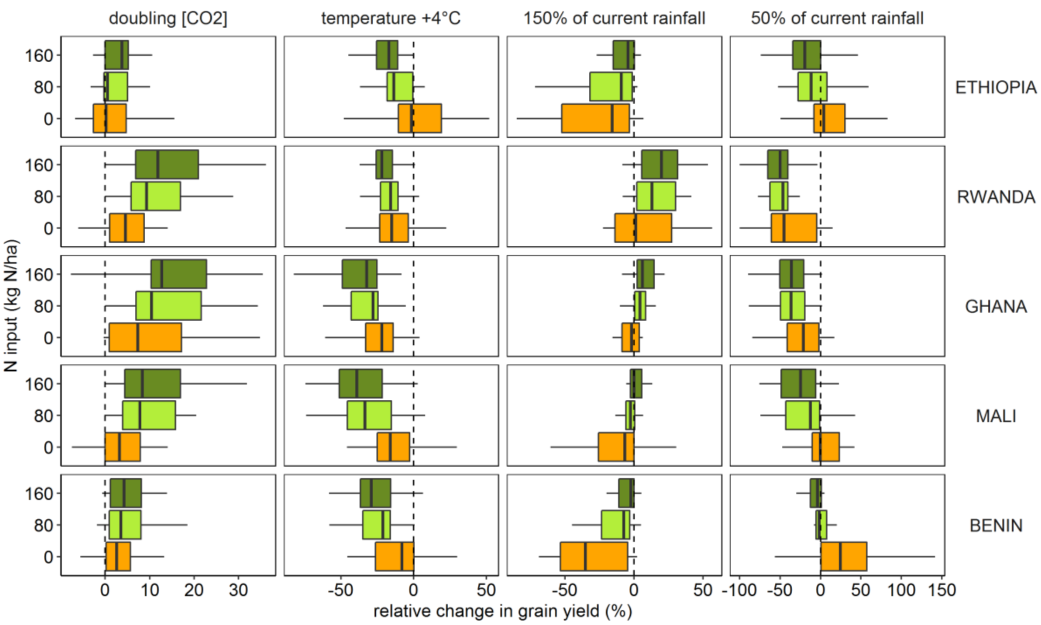
gcb_15261_f1.jpg



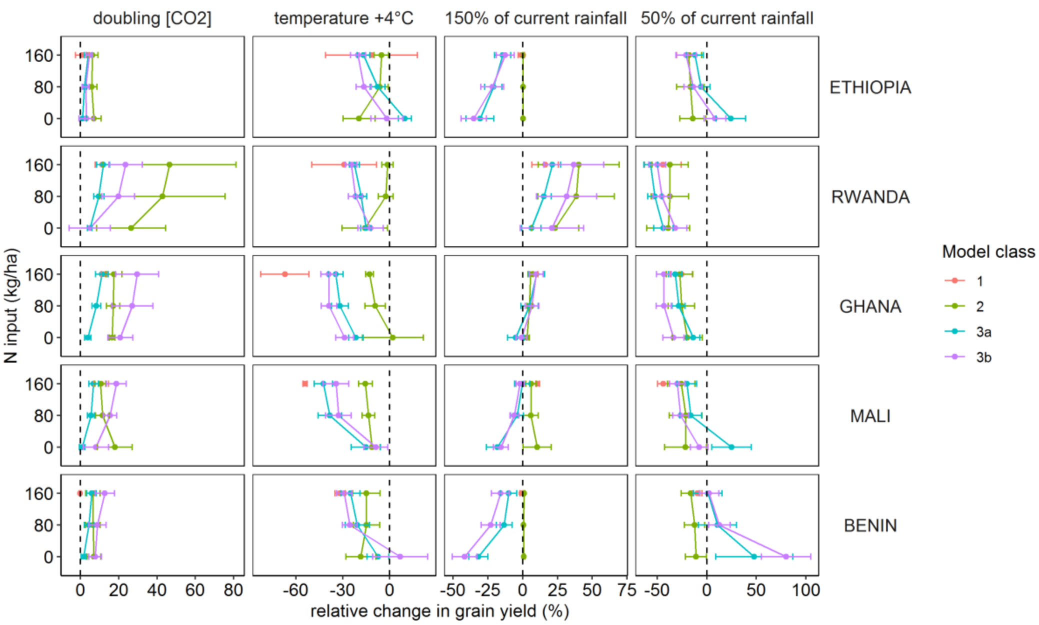
gcb_15261_f2.jpg



gcb_15261_f3.tif



gcb_15261_f4.tif



gcb_15261_f5.tif