Using the anomaly forcing Community Land Model (CLM 4.5) for crop yield projections

Yaqiong Lu[1,2*] and Xianyu Yang[3]
[1]Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 610040, China
[2]National Center for Atmospheric Research, Boulder, CO 80305, USA
[3]Chengdu University of Information Technology, Chengdu, 610225, China
*Corresponding author: Yaqiong Lu, yaqiong@imde.ac.cn, 8602861158015, #.9, Block 4 , Renminnanlu Road, Chengdu 610041, China

Abstract

Crop growth in land surface models normally requires high temporal resolution climate data (3-hourly or 6-hourly), but such high temporal resolution climate data are not provided by many climate model simulations due to expensive storage, which limits modeling choice if there is an interest in a particular climate simulation that only saved monthly outputs. The Community Land Surface Model (CLM) has proposed an alternative approach for utilizing monthly climate outputs as forcing data since version 4.5, and it is called the anomaly forcing CLM. However, such an approach has never been validated for crop yield projections. In our work, we created anomaly forcing datasets for three climate scenarios (1.5 °C warming, 2.0 °C warming, and RCP4.5) and validated crop yields against the standard CLM forcing with the same climate scenarios using 3-hourly data. We found that the anomaly forcing CLM could not produce crop yields identical to the standard CLM due to the different submonthly variations, and crop yields were underestimated by 5-8% across the three scenarios (1.5 °C, 2.0 °C, and RCP4.5) for the global average, and 28-41% of cropland showed significantly different yields. However, the anomaly forcing CLM effectively captured the relative changes between scenarios and over time, as well as regional crop yield variations. We recommend that such an approach be used for qualitative analysis of crop yields when only monthly outputs are available. Our approach can be adopted by other land surface models to expand their capabilities for utilizing monthly climate data.

Key words: Community Land Model; Crop yields; Anomaly forcing

Introduction

Increasing numbers of future climate scenarios exhibit large uncertainties for crop yield projections. Crop yields may increase or decrease depending on which climate projection is used (Lobell et al., 2008; Rosenzweig et al., 2014; Urban et al., 2012). Ensemble future climate projections, such as CMIP5, showed a large range of future climate projections, even for one emission scenario (Knutti and Sedlacek, 2013). Using all future climate projections is not realistic not only because of the computational expense but also because many of these future climate projections only save monthly climate outputs that are not suitable for crop models that require high temporal resolution forcing data. Some standalone process-based crop models run in daily time steps, and some crop models embedded in land surface models need at least 6-hour climate data as the forcing data to represent diurnal cycles. Only a small portion of the CMIP5

47   (Coupled Model Intercomparison Project 5) simulations (<25%) can be used as the forcing data
48   for crop models, leaving little room for crop modelers to choose a particular climate model
49   projection that is of interest.
50
51   The Community Land Model (CLM) (Oleson et al., 2013) is a state-of-the-art land surface model
52   that simulates biogeophysical (radiation transfer, vegetation-soil-hydrology, surface energy
53   fluxes, etc.) and biogeochemical (soil carbon and nitrogen cycle, vegetation photosynthesis,
54   dynamic vegetation growth, etc.) processes. CLM is the default land model in the Community
55   Earth System Model (CESM) (Hurrell et al., 2013), and it can be run either online coupled with
56   the rest of CESM (atmosphere and ocean) or offline (the land model only, forced with climate
57   datasets) for multiple spatial extents (site, regional, and global) and at different resolutions. The
58   crop model derived from AgroIBIS (Kucharik, 2003) was introduced to CLM4.0 by Levis et al.
59   (2012), and it is responsible for crop growth phenology (temperature determined), carbon
60   allocation algorithms, and crop management (e.g., irrigation). The crop model in CLM runs
61   when the soil biogeochemical component is active, and it was tested with the CLM-CN in
62   version 4.0 and tested with CLM-BGC in version 4.5, where CLM-CN and CLM-BGC are
63   officially supported soil biogeochemical components in CLM4.0 and CLM4.5 respectively.
64   Since their introduction, crop models in the CLM have been developed to represent more crop
65   types and processes, such as soybean nitrogen fixation (Drewniak et al., 2013), ozone impacts on
66   yields (Lombardozzi et al., 2015), winter wheat growth responses to cold hazards (Lu et al.,
67   2017), and maize growth responses to heat stress (Peng et al., 2018). CLM simulates nine crop
68   types, accounting for 54% of global total crop production (other production is represented by the
69   most similar crop type): maize, soybean, spring wheat, winter wheat, cotton, rice, sugarcane,
70   tropical maize, and tropical soybean. In this study, we used CLM version 4.5 (Oleson et al.,
71   2013).
72
73   Since version 4.5, CLM offers a built-in function that indirectly uses monthly climate outputs as
74   the forcing data, and is called the anomaly forcing CLM (Lawrence et al., 2015). Anomaly
75   forcing CLM reconstructs new subdaily forcing data by applying the precalculated future
76   monthly anomaly signals to user-defined historical subdaily forcing data, referred to as the
77   reference data. The future monthly anomaly signals are calculated by the future monthly climate
78   outputs and by use of historical monthly outputs. The choice of reference data is arbitrary. Any
79   existing subdaily forcing data (e.g., CRUNCEP (Viovy, 2018), QIAN (Qian et al., 2006)) for
80   CLM can be used as the reference data. The historical monthly outputs are recommended to be
81   multiyear averaged to represent the historical means and avoid affecting the monthly anomaly
82   signal by rare, extreme events in a particular year. Such an arbitrary choice is because the goal of
83   the original anomaly forcing CLM is not to reconstruct future forcing that is identical to the
84   actual future forcing when the high temporal resolution data were saved. Rather, the original goal
85   of the anomaly forcing CLM is to understand the influences due to the anomaly signal by
86   comparing the simulation with the anomaly forcing CLM to the simulation run with the reference
87   data. The differences between the two simulations are due to the anomaly signals.
88
89   In our study, we modified the anomaly forcing CLM to fit our goals to understand whether we
90   could simply use the anomaly forcing CLM for crop yield projections when only monthly
91   climate data were available. We carefully chose the historical monthly data and the reference
92   data so that the reconstructed future anomaly forcing had nearly identical monthly means as the

Deleted: compset

Deleted: compset

Deleted: .

96   desired subdaily future forcing, but we used different submonthly variations. We created
97   anomaly forcing datasets for three future scenarios (1.5 °C warming, 2.0 °C warming, and
98   RCP4.5) for 2006-2075 for which both the subdaily and monthly climate outputs were available
99   from three CESM simulations. With the three paired CLM simulations, we validated the
100 anomaly forcing CLM by comparing it to the standard CLM.
101
102 Methods
103
104 The original anomaly forcing CLM has been available since CLM4.5. This approach reconstructs
105 the subdaily (3-hourly or 6-hourly) forcing data by applying the monthly anomaly signal to user
106 selected subdaily reference data; therefore, it indirectly uses the monthly atmospheric outputs as
107 the forcing data for CLM. This approach does not change any of the scientific code in CLM; it
108 only adds code that reads the monthly anomaly signals and automatically applies these to the
109 reference data while the CLM is running. There were two monthly anomaly signals for RCP4.5
110 and RCP8.5 that were generated using the CESM future projections and were ready for use. It is
111 the user's choice to select which subdaily reference (e.g., CRUNCEP or CLMQIAN) and which
112 years to use. By simply modifying user_nl_cpl namelist and adding data streams of the anomaly
113 forcing variables (see the appendix for the detailed usage), the anomaly forcing CLM will
114 automatically read the monthly anomaly signal and apply the signal to each time step of the
115 reference data within a month. When the reference data period is less than the anomaly signal
116 period, the anomaly forcing CLM will cycle the same reference data until the simulation is
117 complete. Because the different selections of reference data can generate different forcings, even
118 with the same monthly anomaly signals, one should not use the simulation from the anomaly
119 forcing CLM to represent the actual simulation. Rather, the original goal of the anomaly forcing
120 CLM is to compare the simulation with the anomaly forcing and simulation with the reference
121 forcing data to understand the effects of the monthly anomaly signals on land surface variables.
122
123 The goal of this work is to test how well crop yield projections from the anomaly forcing CLM
124 compare to the projections from the standard forcing CLM, given that anomaly forcing has the
125 same monthly average as standard forcing. We selected three future scenarios for CESM
126 simulations that saved both monthly outputs and 3-hourly outputs, where the 3-hourly outputs
127 were directly used in the standard forcing CLM, and the monthly outputs were indirectly used in
128 the anomaly forcing CLM. We calculated the anomaly forcing signals using the monthly CESM
129 outputs and the monthly average of reference data, so that when applying the anomaly signals to
130 the reference data, it is expected to generate identical monthly means as does regular forcing.
131 However, due to a limit in calculations of precipitation anomalies (precipitation anomaly ratio
132 less than 5 times) and how the CLM treats snow and rainfall, the anomaly forcing CLM did not
133 show identical snow and rainfall monthly averages and introduced bias in the crop yield
134 simulations (see the results section).
135
136 Table 1. A summary of the original anomaly forcing CLM and the modifications in this work

	Original anomaly forcing CLM	Modifications in this work
3 h/6 h reference data	User choice	6 h Community Atmosphere Model outputs from one historical low warming

		ensemble simulation 1996-2005
Monthly anomaly signals	Existing for RCP4.5 and RCP8.5	• Anomalies between future scenarios and monthly means of reference data • Three future scenarios: 1.5 °C, 2.0 °C, and RCP4.5 • Each scenario had monthly outputs and 3 h outputs
Goals	Climate impact due to anomaly signals when comparing the anomaly run with the reference run	Given that anomaly forcing has the same monthly mean as the standard CLM forcing, can we use it for crop yield projections?

139
140 We randomly chose the 6-hourly reference data (1996-2005) from one of the 11 historical low
141 warming ensemble CESM simulations. Additionally, we selected three CESM future simulations
142 for the 1.5 °C warming, 2.0 °C warming, and RCP4.5 scenarios, where all the three simulations
143 saved both the monthly outputs and the 3-hourly outputs. We then calculated the monthly
144 anomaly signal at each grid cell for each scenario (1.5, 2.0, and RCP45) from 2006-2075. The
145 monthly anomaly signals are differences for temperature, specific humidity, wind, and air
146 pressure and are ratios for solar radiation and precipitation between the monthly outputs of each
147 scenario and the 1996-2005 averaged monthly values of the reference data. The anomaly forcing
148 signal has both spatial and monthly variations. When running the anomaly forcing simulation for
149 2006-2070, CLM repeatedly uses the 10-year reference period and applies the anomaly signal of
150 a month to all subdaily reference forcing in this month. For example, an anomaly forcing
151 simulation for 2006 January uses the 1996 January reference data plus or multiplied by (if the
152 anomaly signal is a ratio) the 2006 January anomaly signal. If the 2006 January temperature
153 anomaly is 1 K for a grid cell, then all 1996 January reference data will be increased by 1 K for
154 the grid cell.

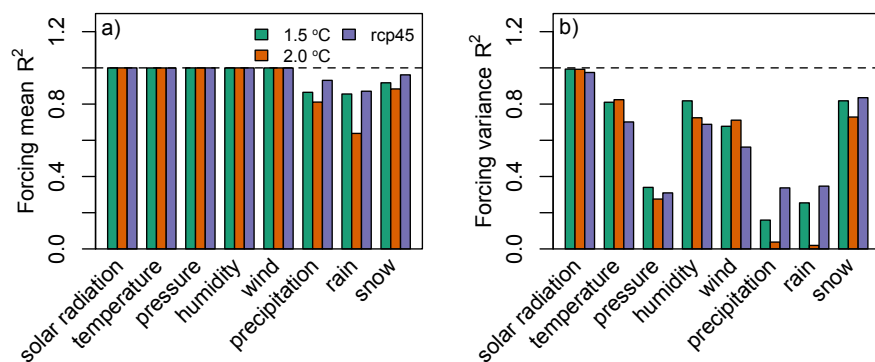
155
156 The monthly anomaly signal is calculated at each grid cell (i,j). For temperature, pressure, wind,
157 and humidity, the anomaly signal is the difference between the future monthly data and the
158 historical monthly average (equation 1). For solar radiation, longwave radiation, and
159 precipitation, the anomaly signal is the ratio between the future monthly data and the historical
160 monthly average (equation 2). We set the maximum ratio for precipitation to 5 to avoid
161 unrealistic extreme precipitation, which also introduced biases in precipitation (discussed in the
162 discussion section).
163
164
$$af_{i,j,m} = fut_{i,j,m} - hist_{i,j,m} \qquad (1)$$
165

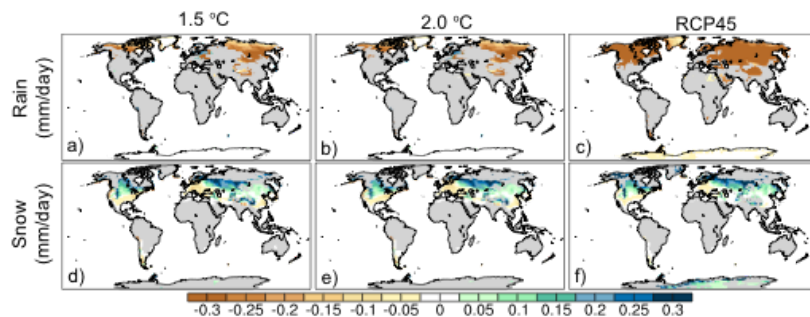
171 $$af_{i,j,m} = fut_{i,j,m}/hist_{i,j,m} \quad (2)$$

172 Where $af_{i,j,m}$ is anomaly forcing signal at a location i and j in a month m, $fut_{i,j,m}$ is the averaged
173 future value and $hist_{i,j,m}$ is the averaged historical value at a location i and j in a month m.
174
175 We set up global CLM crop simulations (compset CLM45BGCCROP) at 1.9 by 2.5 in latitude
176 and longitude, respectively, using the anomaly forcing CLM and the regular forcing CLM for the
177 1.5 °C warming, 2.0 °C warming, and RCP4.5 scenarios. All simulations used the default
178 nitrogen fertilization rates and a constant CO_2 level of 359.8 ppm. For each scenario, we validate
179 the crop yield in the anomaly forcing CLM to the regular forcing CLM to determine if we can
180 use the anomaly forcing CLM for future crop yield projections. We also studied whether the
181 anomaly forcing CLM has a similar crop growth response to transient CO_2 and nitrogen
182 fertilization. The transient CO_2 and nitrogen fertilization did not add extra computational cost
183 compared to the constant CO_2 and nitrogen fertilization simulation. However, due to our limited
184 computational resources could not afford more experiments, we only tested such responses for
185 the RCP4.5 scenario. The transient CO_2 levels in the RCP45 scenario gradually increased from
186 379 ppm in 2006 to 530 ppm in 2070. To test the nitrogen fertilization effects, we simply added a
187 zero nitrogen fertilization simulation here. For the crop yield analysis, we aggregated the
188 individual crop yield into an integrated crop yield by area weighted mean based on the crop area
189 map MAPSMAP (https://www.mapspam.info/) 2005 crop area. The regional crop yield was
190 simply the regional averaged crop yield at 9 regions defined in Ren et al., (2018).
191
192 We adopted the two-sample Kolmogorov-Smirnov test (KS test) to test the statistical
193 significance of differences between the anomaly forcing CLM and the standard CLM for
194 atmospheric forcing data and yield. We used the KS test because some variables at some grid
195 cells did not necessarily follow normal distributions. The KS test is a nonparametric test that
196 detects differences in the empirical probability distributions between two samples, and the two
197 samples do not need to have normal distributions (Justel et al., 1997; Marozzi, 2013). When
198 repeated using the ten-year reference data, we expected that the ten year averaged monthly
199 anomaly forcing would show no significant differences from the regular forcing. Thus, for the
200 atmospheric forcing data, we tested probability distribution differences between anomaly forcing
201 and regular forcing for every ten-year averaged monthly dataset (sample size was 7x12=84). For
202 crop yields, we used the every ten-year averaged annual yields (sample size was 7). We used
203 linear regression coefficent (R^2), bias (equation 3), percentage differences (equation 4) in our
204 evaluations.
205
206 $$bias = CLM_{anomaly\ forcing} - CLM_{standard} \quad (3)$$
207 $$\%differences = 100 * (\frac{CLM_{anomaly\ forcing}}{CLM_{standard}} - 1) \quad (4)$$
208
209
210 Results
211
212 We aimed to generate an anomaly forcing that produced identical monthly averages as its
213 counterpart regular forcing (the desirable 3-hourly forcing data for CLM) but with different
214 submonthly variations. All atmospheric forcing variables achieved this goal except for
215 precipitation and its liquid and ice components, rain and snow. The linear regression coefficients

219 (R^2) between anomaly forcing and standard forcing for the monthly means of incoming solar
220 radiation, bottom layer atmosphere temperatures (sigma vertical coordinate, σ=0.9925) ,
221 pressures, humidities, and winds all showed R^2 values above 0.99, and there were also no
222 significant differences for these variables for all grid cells. However, for rain and snow, the R^2
223 values were 0.63-0.87 and 0.88-0.96 across the three scenarios, respectively (Figure 1a).
224 Statistically significant differences were also found for rain and snow in many regions in the
225 Northern Hemisphere (Figure 2). We used monthly variances as a measure of the submonthly
226 variations. We calculated the variation for twelve months in each decade, so we have 7 decades
227 and 12 months variance and the sample size is 84 when setting up the regression. R^2 for
228 variances of forcing were low for most variables except for incoming solar radiation (Figure 1b).
229 Such lower R^2 values indicated that anomaly forcing could not represent the submonthly
230 variations as well as the regular forcing.
231

232 There were two error sources for precipitation. First, there was overall average lower
233 precipitation in the anomaly forcing by 0.02 mm/day, 0.03 mm/day, and 0.2 mm/day in the
234 1.5 °C, 2.0 °C, and RCP45 scenarios, respectively. Such slightly lower precipitation was because
235 we set the maximum precipitation anomaly ratio to 5 to avoid unrealistically extreme
236 precipitation levels. Ratio 5 was suggested by NCAR scientists David Lawrence and Sean
237 Swenson, who are core developers of CLM and wrote the initial anomaly forcing code in CLM.
238 Most of unrealistic extreme precipitation ratio are actually due to the nearly zero historical
239 precipitation (the denominator of equation 2). The cap for the precipitation anomaly ratio is use
240 to avoid such situation. Second, the CLM used the temperature in each time step to determine if
241 the given precipitation was rain or snow. Precipitation was rain when temperature was above
242 273.15 K, otherwise it was snow. Therefore, the different submonthly variations in temperature
243 resulted in different submonthly variations for snow and rain. Due to this problem, the lower
244 precipitation did not evenly distribute to the rain and snow bias, for which rain was
245 underestimated by 0.08-0.3 mm/day, and snow was overestimated by 0.06-0.11 mm/day across
246 the three scenarios. The significantly different regions were mainly in the Northern Hemisphere
247 and the Antarctic, and most regions in the Southern Hemisphere did not show significant
248 differences in rain or snow. How the rain and snow biases affected yield projections will be
249 discussed.

250

251
252 Figure 1. Linear regression coefficients (R^2) between a) decade-averaged monthly mean (sample
253 size =12 months x 7 decades=84) between anomaly forcing and regular forcing and b) every ten
254 year-averaged monthly variance between anomaly forcing and regular forcing.

255

256
257
258 Figure 2. 70-year averaged differences between anomaly forcing and regular forcing for rain (a-
259 c) and snow (d-f) for the 1.5°C, 2.0 °C, and RCP4.5 scenarios. All differences shown here are
260 statistically significant differences tested by the Kolmogorov-Smirnov test with a sample size of
261 84. The gray areas are regions that did not show significant differences.
262
263 When compared to crop yield simulations in the standard CLM, the anomaly forcing CLM
264 underestimated crop yields by 5-8% across the three scenarios for the global average, and 28-
265 41% of cropland showed statistically significant differences in yields. The rainfed crop yield
266 differences across the three scenarios showed largely similar spatial distributions: overestimation
267 in the northern US and Europe and underestimation in the Southern Hemisphere and in East Asia

(Figure 3d-f). The overestimated rainfed crop yield (mainly for maize and wheat) in the anomaly forcing CLM is due to higher water availability in these regions, which is a result of higher snow in the anomaly forcing CLM. For irrigated crops, such overestimations in the northern US and Europe disappear (Figure 3g-i) because sufficient irrigation was added to the irrigated soil column in the standard CLM, which removed the plant water stress that was seen for rainfed crops. However, the underestimations in the Southern Hemisphere and East Asia were persistent, because water availability does not cause yield differences for irrigated crops; we suspect such underestimations were caused by the other error in forcing data: the different submonthly variations in the forcing data.

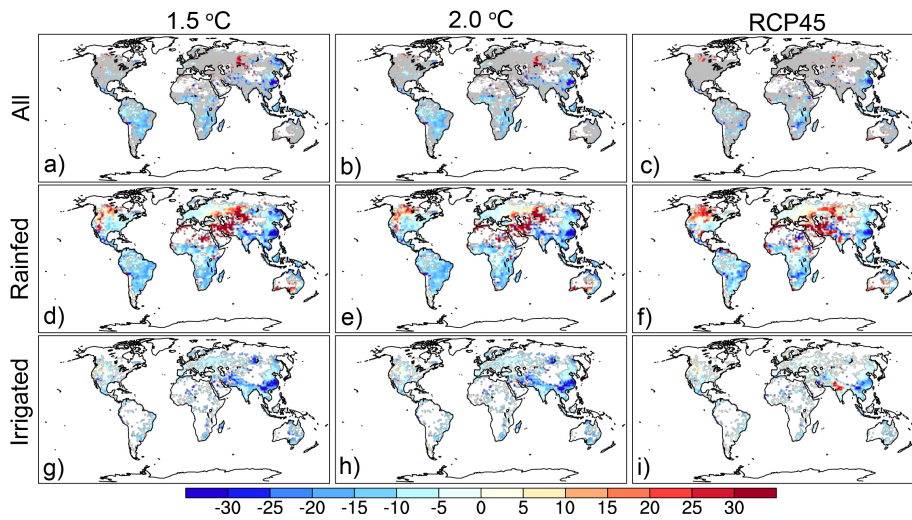
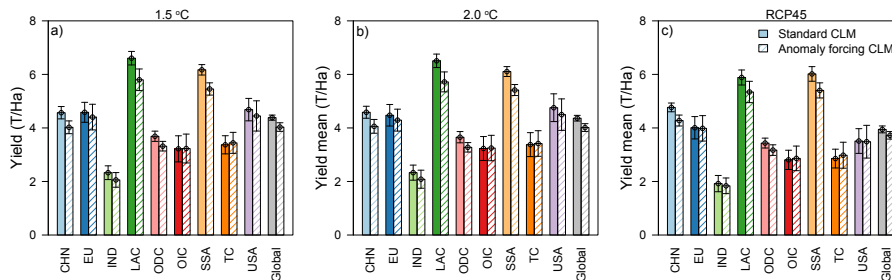



Figure 3. The percentage differences of 70-year integrated yields between the anomaly forcing CLM and the standard CLM for all crops (a-c), rainfed crops (d-f), and irrigated crops (g-i) for the 1.5 °C, 2.0 °C, and RCP45 scenarios. The white regions are where no crops grow based on the historical crop map in 2005 (MAPSPAM 2005; https://www.mapspam.info/). For plots a-c, we showed only the significant differences as determined by the by Kolmogorov-Smirnov test with a sample size of 7. The regions with insignificant differences are masked as gray in a-c. For plots d-i, we did not mask the insignificant differences to show an overall bias.

297
298 Figure 4. Regional comparisons of the 70-year integrated mean yields and yield standard
299 deviations between the anomaly forcing CLM and the standard CLM. The error bars indicate 70-
300 year yield standard deviations. CHN: China; EU: European Union; IND: India; LAC: Latin
301 America; ODC: Other Developing Countries; OIC: Other Industrialized Countries; SSA: Sub-
302 Saharan Africa; TC: Transition Countries; USA: United States

303 The global 70-year averaged yields ± standard deviation in the standard CLM (Ren et al., 2018)
304 and in the anomaly forcing CLM are 4.38 ± 0.09 and 4.03 ± 0.16 t/ha, respectively, in the 1.5 °C
305 scenario, 4.36 ± 0.11 and 4.01 ± 0.14 t/ha, respectively, in the 2.0 °C scenario, 3.95 ±0.13 and
306 3.72 ± 0.14, respectively, in the RCP45 scenario (Figure 4). The anomaly forcing CLM captured
307 the regional yield variations. Latin America (LAC) showed the highest yield while India (IND)
308 showed the lowest yields for both the anomaly forcing CLM and the standard CLM across the
309 three scenarios.
310
311
312 Although the crop yields were underestimated, the anomaly forcing CLM could qualitatively
313 represent the spatial yield differences between two climate scenarios. Comparing 2.0 °C to
314 1.5 °C, there was a 4-8% yield increase in the northern U.S. and a 0-4% yield decrease in (Figure
315 5a) in the southeast U.S. When comparing the RCP45 to the 1.5 °C scenario, crop yields in the
316 U.S. were largely reduced (up to 50%). The anomaly forcing CLM clearly captured these yield
317 differences (Figure 5b and 5d).

Deleted:

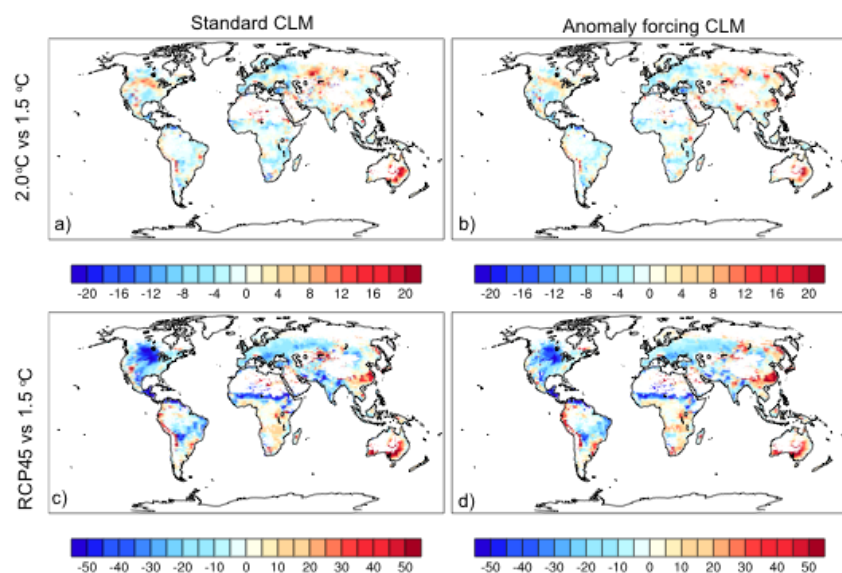



Figure 5. The percentage of 70-year integrated yield differences between 2.0 °C and 1.5 °C (top panel) RCP45 to 1.5 (bottom panel) in the standard CLM and the anomaly forcing CLM

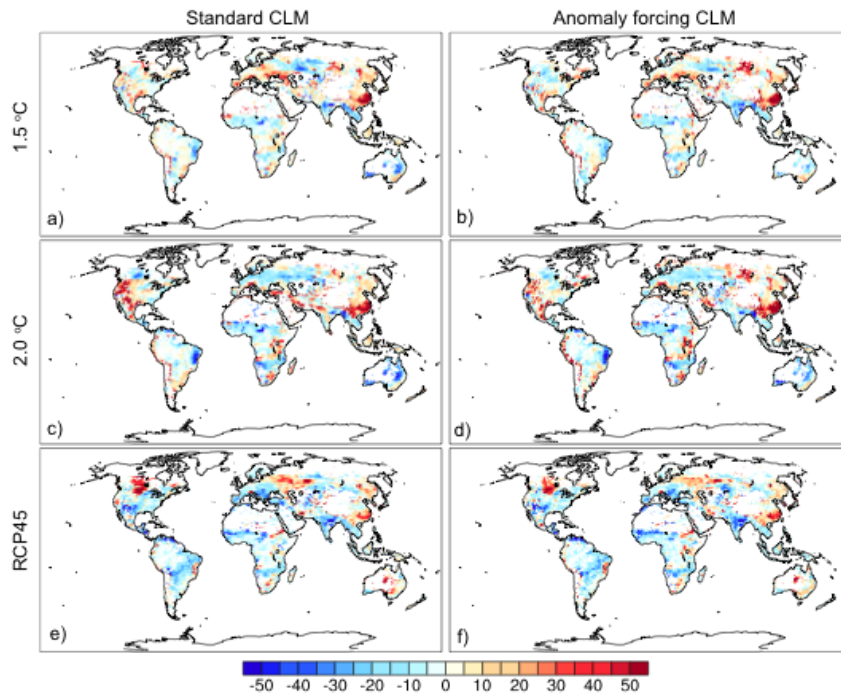
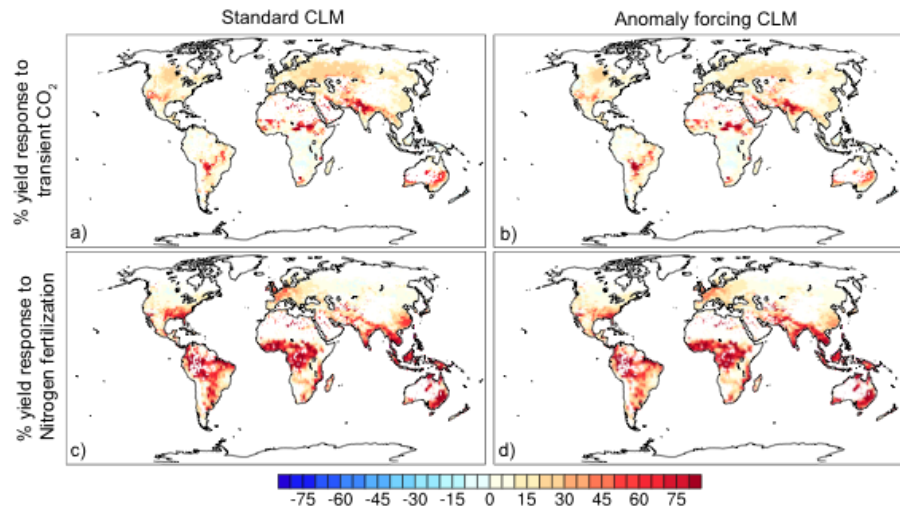



Figure 6. The percentage yield difference from 2006-2015 to 2066-2075 in the standard CLM
and anomaly forcing CLM across the three scenarios

The anomaly forcing CLM also captured yield changes over time for each climate scenario. The
three scenarios showed some similarities in yield changes from 2006-2015 to 2066-2075. For
example, crop yields increased in Southeast China and decreased in Sub-Saharan Africa. There
were also yield changes that were unique to each scenario that were also found in the anomaly
forcing CLM. For example, crop yields increased in Europe for the 1.5 °C scenario (Figure 6a-
b), while they decreased in Europe for the 2.0 °C and RCP45 scenarios (Figure 6c-f), and crop
yields declined in the U.S. for the RCP45 scenario (Figure 6e-f) while they increased for the
1.5 °C and 2.0 °C scenarios (Figure 6 a-d).

All simulations in the above evaluations adopted a constant CO_2 level (359.8 ppm) and crop
types dependent fixed nitrogen fertilization (25-500 kg N/ ha), so whether the anomaly forcing
CLM simulated a similar or different crop growth response to CO_2 or nitrogen fertilization is
unknown. Due to limited computational resources, we tested crop responses to transient CO_2 and
nitrogen fertilization only for the RCP45 scenario and assumed that the other scenarios would
show the same differences as the RCP45 scenario. The transient CO_2 in the RCP45 scenario
gradually increased from 379 ppm in 2006 to 530 ppm in 2075. To test the effects of nitrogen

348 fertilization, we simply added a zero nitrogen fertilization simulation. Although all grid cells had
349 the same amounts of CO_2 increase in a given year (no spatial variation), crop yields had spatial
350 variations in response to transient CO_2. Most regions showed a 5-10% yield increase, but some
351 regions showed much higher yield increases, such as northern India, the southern edge of the
352 Sahara, and Australia (Figure 7a). Such crop yield responses to transient CO_2 spatial patterns
353 were also captured by the anomaly forcing CLM (Figure 7b). Similar for the crop yield responses
354 to nitrogen fertilization, the anomaly forcing CLM simulated crop yield increase spatial patterns
355 (Figure 7c-d), in which the Southern Hemisphere and Asia had greater yield increases in
356 response to nitrogen fertilization.
357
358

359
360
361   Figure 7. 70-year averaged integrated crop yield response to transient CO_2 and to no nitrogen
362   fertilization in the anomaly forcing CLM (a and b) and in the standard CLM (c and d) for the
363   RCP45 scenario.
364
365 Discussion
366
367 In this work, we created anomaly forcing datasets for three future climate scenarios, and we
368 validated the crop yields in the anomaly forcing CLM by comparison with the crop yields in the
369 standard CLM. The differences between the anomaly forcing CLM and standard CLM were due
370 only to differences in forcing data, for which the standard CLM used regular forcing (three-
371 hourly forcing) and the anomaly forcing CLM used anomaly forcing. We found that the anomaly
372 forcing CLM underestimated crop yields but identified the regional yield variations, as well as
373 yield differences between two climate scenarios and yield changes over time. The anomaly
374 forcing CLM could not generate the exact same crop yields as the standard CLM due to errors in

375 precipitation and in the submonthly variations. However, it could be used for qualitative analysis
376 of relative crop yield changes among different scenarios and over time.
377
378 The overall underestimation of crop yields may be due to differences in phenology that resulted
379 from different submonthly variations. Some of the low yields in the anomaly forcing CLM may
380 be explained by shorter grain fill periods. For example, the lower rice yields in southeast China
381 are due to a 5-10 day shorter grain fill period in the anomaly forcing CLM (Figure S1;a-c); maize
382 and soybean in the Southern Hemisphere also showed a 1-5 day shorter grain fill period that may
383 account for the lower yields (Figure S1; d-i). In addition to the low yields, the anomaly forcing
384 CLM also simulated lower GPP and LAI compared to the standard CLM (Figure S2; a1-b3), and
385 the spatial distributions of GPP and LAI differences were very similar to the yield differences.
386
387 Some regions in the Northern Hemisphere showed higher rainfed crop yields in the anomaly
388 forcing CLM, which is due to higher soil moistures at planting that resulted from higher snow
389 levels in the Northern Hemisphere. Crop growth in CLM is very sensitive to the soil moisture at
390 planting, and higher soil moisture (Figure S2; c1-c3) results in unstressed crop growth and hence
391 produces higher yields. When adequate irrigation is applied, both the anomaly forcing and the
392 standard CLM models have sufficient water for crop growth, and the overestimations
393 disappeared. Therefore, the anomaly forcing may not be appropriate for estimating the actual
394 future irrigation demands but is able to distinguish the relative differences in irrigation demand
395 across different climate scenarios.
396
397 The energy fluxes in the anomaly forcing CLM and in the standard CLM were different due to
398 different crop growth rates and differences in forcing data. The higher snow cover in the
399 Northern Hemisphere creates higher albedo and lowers absorbed solar radiation and hence lower
400 surface energy fluxes. The higher LAI increased the summer latent heat flux up to 5 W.m^{-2}
401 (Figure S3), while the annual latent heat flux showed 5-10 W.m^{-2} (Figure S2; d1-d3) lower
402 values in the anomaly forcing CLM due to the lower net radiation. In the Southern Hemisphere,
403 lower LAI (Figure S2; a1-a3) resulted in lower latent heat fluxes (Figure S2; d1-d3) and higher
404 sensible heat fluxes (Figure S2; e1-e3).
405
406 The regional yield comparisons indicate that the anomaly forcing CLM effectively captured
407 regional yield variations but with slightly lower yield biases. We want to point out that the very
408 high crop yields in Latin America and in Sub-Saharan Africa, and the very low crop yields in
409 India in both the anomaly forcing CLM and the standard CLM approaches are not realistic when
410 compared to the UNFAO yields (http://www.fao.org/statistics/en/). Such biases in the CLM have
411 been discussed by Levis et al. (2018), and the low yields in India are due to incorrect crop
412 phenology when crops entered the grain fill during the dry season. The high yields in Latin
413 American and in Sub-Saharan Africa were due to the nitrogen fertilization amounts based on US
414 levels, which are too high for these regions.
415
416 The crop model in the most recent version of CLM5.0 includes new features as reported in
417 Lombardozzi et al., (2020). For example CLM5.0 uses time-varying spatial distributions of
418 major crop types and has updated fertilization and irrigation schemes. These updates of crop
419 model in CLM5.0 may improve the crop yield simulations for both standard CLM and anomaly
420 forcing CLM compared to crop yield in reality. The anomaly forcing method in CLM5.0 remains

Deleted: is due

Deleted: not shown),

Deleted:

Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1

424 unchanged so we speculate the bias due to anomaly forcing may still exist in CLM5.0. For
425 example, CLM5.0 uses the same threshold to differ rain and snow, so the bias due to higher snow
426 cover in the Northern Hemisphere may still exists in CLM5.0. However, how will the magnitude
427 of the bias change is unclear. We suggest that the anomaly forcing of CLM5.0 to be tested if the
428 research interest is in absolute yield or in qualitative difference.
429
430 Our approach can be adopted by other land surface models to expand their capabilities for
431 utilizing monthly climate data. The source code of the anomaly forcing CLM is available at the
432 repository website Zenodo https://doi.org/10.5281/zenodo.3900671. The path is
433 post4.5crop_slevis/models/lnd/clm/src/cpl/ lnd_import_export.F90 when unzip
434 post4.5crop_slevis_codeforGMD.tar.gz. The Fortran code could be transplanted to other land
435 surface models which use NetCDF format atmospheric forcing.
436
437 Conclusions
438
439 The Community Land Surface model offers an alternative way in utilize the monthly climate as
440 the forcing data. Such an approach could expand user choice of forcing data when high temporal
441 resolution climate data are not available. In this work, we created anomaly forcing data for three
442 climate scenarios (1.5 °C warming, 2.0 °C warming, and RCP4.5) and validated crop yield
443 projections in the anomaly forcing CLM against the standard CLM. The anomaly forcing CLM
444 underestimated crop yields by 5-8%, which was largely due to the differences in phenology and
445 photosynthesis that resulted from the different submonthly variations. How CLM treated
446 precipitation as rain or snow also introduced biases in crop yields and in the energy flux
447 simulations. Although the anomaly forcing CLM could not generate crop yields identical to the
448 standard CLM, it could be used for qualitative analysis of crop yield changes across various
449 scenarios over time.
450
451 Code availability
452
453 The CLM source code used in our study is available at repository website Zenodo:
454 https://doi.org/10.5281/zenodo.3900671
455
456 Author contribution
457
458 Yaqiong Lu designed and performed the simulations. Yaqiong Lu and Xianyu Yang analyzed the
459 results and wrote the manuscript.
460
461 Acknowledgments
462

463 We thank the tropical editor Christoph Müller and two anonymous reviewers for their great
464 comments that largely improved our manuscript. We also thank Sean Swenson and David
465 Lawrence for instruction of using the anomaly forcing approach. This work was supported by the
466 National Science Foundation under Grant Number AGS-1243095 and the National Natural
467 Science Foundation of China (No. 41975135). We would like to acknowledge high-performance
468 computing support from Yellowstone (ark:/85065/d7wd3xhc), provided by NCAR's

470   Computational and Information Systems Laboratory, sponsored by the National Science
471   Foundation.
472
473
474   Appendix: a user guide for using anomaly forcing CLM
475
476   Running the anomaly forcing CLM is similar to the standard CLM but with several additional
477   steps. First, the monthly anomaly data are prepared as described in the method section. Then, the
478   user needs to modify user_nl_cpl and user_nl_datm to specify which forcing variables to add to
479   the anomaly signals. There are seven anomaly forcing variables (Table A2), and the user can
480   specify one, or two, or all variables in the two namelists (user_nl_cpl and user_nl_datm). The
481   final step is to add the corresponding anomaly forcing data streams depending on which anomaly
482   forcing variables were specified in user_nl_cpl and user_nl_datm.
483
484
485   1. Modify user_nl_cpl and user_nl_datm
486
487   The user may add part or all of the following text to user_nl_cpl.
488
489   cplflds_custom = 'Sa_prec_af->a2x', 'Sa_prec_af->x2l','Sa_tbot_af->a2x',
490   'Sa_tbot_af->x2l','Sa_pbot_af->a2x', 'Sa_pbot_af->x2l','Sa_shum_af->a2x',
491   'Sa_shum_af->x2l','Sa_u_af->a2x', 'Sa_u_af->x2l','Sa_v_af->a2x',
492   'Sa_v_af->x2l','Sa_swdn_af->a2x', 'Sa_swdn_af->x2l','Sa_lwdn_af->a2x', 'Sa_lwdn_af->x2l'
493
494   Add part or all of the following text into user_nl_datm:
495
496   anomaly_forcing=
497   'Anomaly.Forcing.Precip','Anomaly.Forcing.Temperature','Anomaly.Forcing.Pressure','Anomaly
498   .Forcing.Humidity','Anomaly.Forcing.Uwind','Anomaly.Forcing.Vwind','Anomaly.Forcing.Short
499   wave','Anomaly.Forcing.Longwave'
500
501   Also attach the anomaly forcing data streams in user_nl_datm:
502
503   streams = "datm.streams.txt.CLMCRUNCEP.Solar 1996 1996 2005",
504   "datm.streams.txt.CLMCRUNCEP.Precip 1996 1996 2005",
505   "datm.streams.txt.CLMCRUNCEP.TPQW 1996 1996 2005",
506   "datm.streams.txt.presaero.clim_2000 1 1 1",
507   "datm.streams.txt.Anomaly.Forcing.Precip 2006 2006 2075",
508   "datm.streams.txt.Anomaly.Forcing.Temperature 2006 2006 2075",
509   "datm.streams.txt.Anomaly.Forcing.Pressure 2006 2006 2075",
510   "datm.streams.txt.Anomaly.Forcing.Humidity 2006 2006 2075",
511   "datm.streams.txt.Anomaly.Forcing.Uwind 2006 2006 2075",
512   "datm.streams.txt.Anomaly.Forcing.Vwind 2006 2006 2075",
513   "datm.streams.txt.Anomaly.Forcing.Shortwave 2006 2006 2075",
514   "datm.streams.txt.Anomaly.Forcing.Longwave 2006 2006 2075",
515   "/glade/p/work/yaqiong/inputdata/atm/datm7/co2.1pt5degC.streams.txt 1901 1901 2075"

516
517 mapalgo = 'bilinear', 'bilinear', 'bilinear', 'bilinear', 'bilinear', 'bilinear', 'bilinear', 'bilinear',
518 'bilinear',
519   'bilinear', 'bilinear', 'bilinear','nn'
520 tintalgo = 'coszen', 'nearest', 'linear', 'linear', 'nearest', 'nearest', 'nearest', 'nearest', 'nearest',
521 'nearest',
522   'nearest', 'nearest','linear'
523
524 Any combination or subset of anomaly forcing variables can be used. For example,
525 cplflds_custom = 'Sa_prec_af->a2x', 'Sa_prec_af->x2l' (in user_nl_cpl)
526 anomaly_forcing='Anomaly.Forcing.Precip' (in user_nl_datm)
527 will only adjust precipitation. The reference data and period are defined in env_run.xml.
528
529 2. Add the anomaly forcing data stream
530 The anomaly forcing data stream is where to specify the data path of the monthly anomaly
531 forcing signal and to tell the code which variable to retrieve. A list of all anomaly forcing data
532 stream file names and the variables in the anomaly forcing data and the code are given in Table
533 2. An example of the content in user_datm.streams.txt.Anomaly.Forcing.Humidity is also
534 attached. The user only needs to add the corresponding variable data streams that are defined in
535 user_nl_cpl.
536
537 Table A2. A list of the anomaly forcing data streams and the corresponding variables in the
538 anomaly forcing data and the code

Data stream file names	Vars in data	Vars in code
user_datm.streams.txt.Anomaly.Forcing.Humidity[1]	huss	shum_af
user_datm.streams.txt.Anomaly.Forcing.Precip	pr	prec_af
user_datm.streams.txt.Anomaly.Forcing.Pressure	ps	pbot_af
user_datm.streams.txt.Anomaly.Forcing.Shortwave	rsds	swdn_af
user_datm.streams.txt.Anomaly.Forcing.Temperature	tas	tbot_af
user_datm.streams.txt.Anomaly.Forcing.Uwind	uas	u_af
user_datm.streams.txt.Anomaly.Forcing.Vwind	vas	v_af
user_datm.streams.txt.Anomaly.Forcing.Longwave	rlds	lwdn_af

539 [1]An example of the content in the data stream was given below:
540 <dataSource>
541   GENERIC
542   </dataSource>
543   <domainInfo>
544   <variableNames>
545   time
546   xc   lon
547   yc   lat
548   area
549   mask
550   </variableNames>
551   <filePath>
552   /glade/p/cesmdata/cseg/inputdata/share/domains

```
553     </filePath>
554     <fileNames>
555       domain.lnd.fv0.9x1.25_gx1v6.090309.nc
556     </fileNames>
557   </domainInfo>
558   <fieldInfo>
559     <variableNames>
560       huss  shum_af
561     </variableNames>
562     <filePath>
563       THE ANOMALY FORCING SIGNAL DATA PATH
564     </filePath>
565     <fileNames>
566       THE ANOMALY FORCING SIGNAL DATA NAME
567     </fileNames>
568     <offset>
569       0
570     </offset>
571   </fieldInfo>
572
573
574   Reference:
575
```

576   Drewniak, B., Song, J., Prell, J., Kotamarthi, V. R., and Jacob, R.: Modeling agriculture in the Community Land Model,
577   Geosci Model Dev, 6, 495-515, 2013.
578   Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J. F., Large, W. G., Lawrence,
579   D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S.,
580   Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S.: The Community Earth System Model
581   A Framework for Collaborative Research, B Am Meteorol Soc, 94, 1339-1360, 2013.
582   Justel, A., Pena, D., and Zamar, R.: A multivariate Kolmogorov-Smirnov test of goodness of fit, Stat Probabil Lett, 35,
583   251-259, 1997.
584   Knutti, R. and Sedlacek, J.: Robustness and uncertainties in the new CMIP5 climate model projections, Nat Clim
585   Change, 3, 369-373, 2013.
586   Kucharik, C. J.: Evaluation of a Process-Based Agro-Ecosystem Model (Agro-IBIS) across the US Corn Belt:
587   Simulations of the Interannual Variability in Maize Yield. In: Earth Interact, 2003.
588   Lawrence, D. M., Koven, C. D., Swenson, S. C., Riley, W. J., and Slater, A. G.: Permafrost thaw and resulting soil
589   moisture changes regulate projected high-latitude CO2 and CH4 emissions, Environ Res Lett, 10, 2015.
590   Levis, S., Badger, A., Drewniak, B., Nevison, C., and Ren, X. L.: CLMcrop yields and water requirements: avoided
591   impacts by choosing RCP 4.5 over 8.5, Climatic Change, 146, 501-515, 2018.
592   Levis, S., Bonan, G. B., Kluzek, E., Thornton, P. E., Jones, A., Sacks, W. J., and Kucharik, C. J.: Interactive Crop
593   Management in the Community Earth System Model (CESM1): Seasonal Influences on Land-Atmosphere Fluxes, J
594   Climate, 25, 4839-4859, 2012.
595   Lobell, D. B., Burke, M. B., Tebaldi, C., Mastrandrea, M. D., Falcon, W. P., and Naylor, R. L.: Prioritizing climate
596   change adaptation needs for food security in 2030, Science, 319, 607-610, 2008.
597   Lombardozzi, D., Levis, S., Bonan, G., Hess, P. G., and Sparks, J. P.: The Influence of Chronic Ozone Exposure on
598   Global Carbon and Water Cycles, J Climate, 28, 292-305, 2015.
599   Lombardozzi, D. L., Lu, Y. Q., Lawrence, P. J., Lawrence, D. M., Swenson, S., Oleson, K. W., Wieder, W. R., and
600   Ainsworth, E. A.: Simulating Agriculture in the Community Land Model Version 5, J Geophys Res-Biogeo, 125, 2020.
601   Lu, Y. Q., Williams, I. N., Bagley, J. E., Torn, M. S., and Kueppers, L. M.: Representing winter wheat in the
602   Community Land Model (version 4.5), Geosci Model Dev, 10, 1873-1888, 2017.
603   Marozzi, M.: Nonparametric Simultaneous Tests for Location and Scale Testing: A Comparison of Several Methods,

Commun Stat-Simul C, 42, 1298-1317, 2013.

Oleson, K., Lawrence, D., Bonan, G., Drewniak, B., Huang, M., Koven, C., Levis, S., Li, F., Riley, W., Subin, Z., Swenson, S., and Thornton, P.: Technical Description of version 4.5 of the Community Land Model (CLM), National Center for Atmospheric Rsearch, Boulder, CO, NCAR/TN-503+STR, 434 pp., 2013.

Peng, B., Guan, K. Y., Chen, M., Lawrence, D. M., Pokhrel, Y., Suyker, A., Arkebauer, T., and Lu, Y. Q.: Improving maize growth processes in the community land model: Implementation and evaluation, Agr Forest Meteorol, 250, 64-89, 2018.

Qian, T., Dai, A., Ternberth, K. E., and Olseon, K. W.: Simulation of Global Land Surface Conditions from 1948 to 2004. Part I: Forcing Data and Evaluations, Journal of Hydrometeorology, 7, 953-975, 2006.

Ren, X., Lu, Y., O'Neill, B. C., and Weitzel, M.: Economic and biophysical impacts on agriculture under 1.5 °C and 2 °C warming, Environ Res Lett, 13, 2018.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Muller, 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., and Jones, J. W.: Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison, P Natl Acad Sci USA, 111, 3268-3273, 2014.

Urban, D., Roberts, M. J., Schlenker, W., and Lobell, D. B.: Projected temperature changes indicate significant increase in interannual variability of U.S. maize yields, Climatic Change, 112, 525-533, 2012.

Viovy, N.: CRUNCEP Version 7 - Atmospheric Forcing Data for the Community Land Model. https://doi.org/10.5065/PZ8F-F017, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, 2018.