

# Supplemental Material

## The GGCMI phase II emulators: global gridded crop model responses to changes in CO<sub>2</sub>, temperature, water, and nitrogen (version 1.0)

James Franke<sup>1,2</sup>, Christoph Müller<sup>3</sup>, Joshua Elliott<sup>2,4</sup>, Alexander Ruane<sup>5</sup>, Jonas Jägermeyr<sup>3,2,4,5</sup>,

Abigail Snyder<sup>6</sup>, Marie Dury<sup>7</sup>, Pete Falloon<sup>8</sup>, Christian Folberth<sup>9</sup>, Louis François<sup>7</sup>, Tobias Hank<sup>10</sup>,

Cesar Izaurralde<sup>11,12</sup>, Ingrid Jacquemin<sup>7</sup>, Curtis Jones<sup>11</sup>, Michelle Li<sup>2,13</sup>, Wenfeng Liu<sup>14,15</sup>, Stefan Olin<sup>16</sup>, Meridell Phillips<sup>5,17</sup>, Thomas Pugh<sup>18,19</sup>, Ashwan Reddy<sup>11</sup>, Karina Williams<sup>8</sup>, Ziwei Wang<sup>1,2</sup>, Florian Zabel<sup>10</sup>, and Elisabeth Moyer<sup>1,2</sup>

1. Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA

2. Center for Robust Decision-making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, IL, USA

3. Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Potsdam, Germany

4. Department of Computer Science, University of Chicago, Chicago, IL, USA

5. NASA Goddard Institute for Space Studies, New York, NY, United States

6. Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA

7. Unité de Modélisation du Climat et des Cycles Biogéochimiques, UR SPHERES, Institut d’Astrophysique et de Géophysique, University of Liège, Belgium

8. Met Office Hadley Centre, Exeter, United Kingdom

9. Ecosystem Services and Management Program, International Institute for Applied Systems Analysis, Laxenburg, Austria

10. Department of Geography, Ludwig-Maximilians-Universität, Munich, Germany

11. Department of Geographical Sciences, University of Maryland, College Park, MD, USA

12. Texas AgriLife Research and Extension, Texas A&M University, Temple, TX, USA

13. Department of Statistics, University of Chicago, Chicago, IL, USA

14. EAWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

15. Laboratoire des Sciences du Climat et de l’Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France.

16. Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

17. Earth Institute Center for Climate Systems Research, Columbia University, New York, NY, USA

18. School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK.

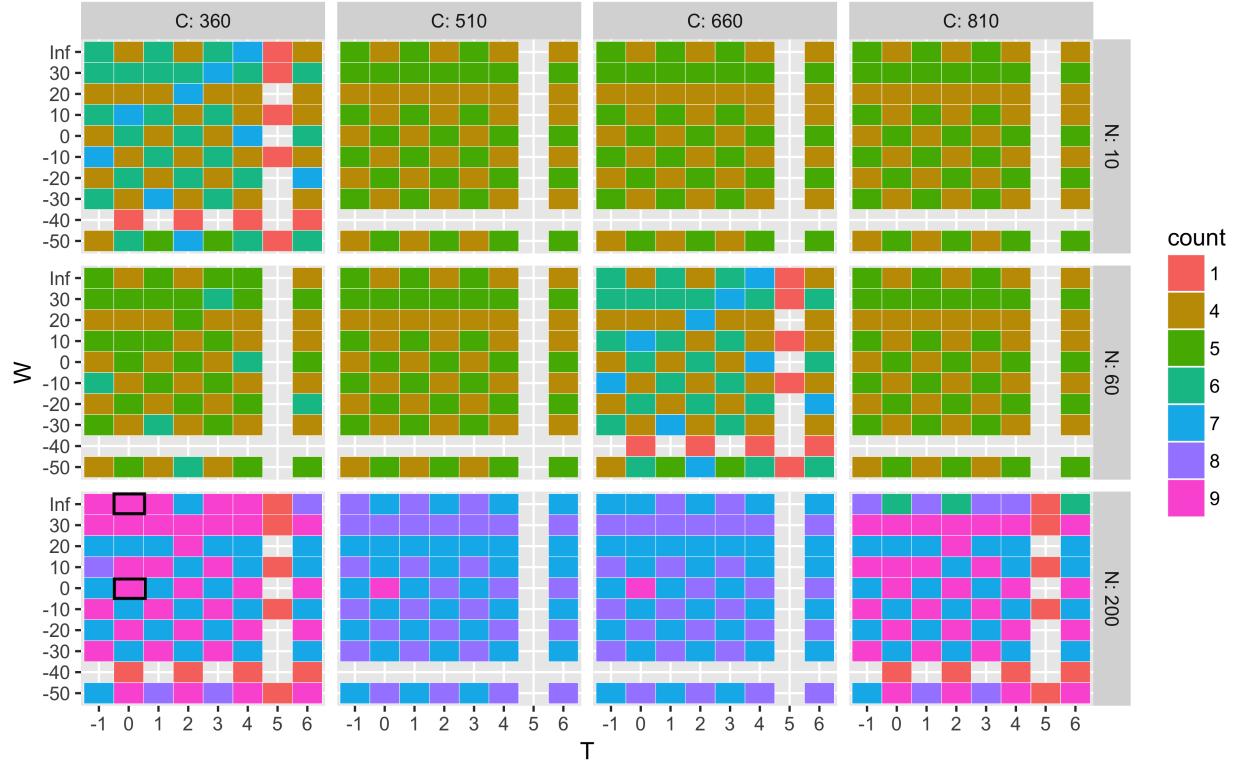
19. Birmingham Institute of Forest Research, University of Birmingham, Birmingham, UK.

## Contents

<b>1 Experiment simulation sampling in variable space</b>	<b>3</b>
<b>2 Cultivated areas</b>	<b>4</b>
<b>3 Variability changes in future climate projections</b>	<b>5</b>
<b>4 Yield response for A1 (growing season adaptation) simulations</b>	<b>6</b>
<b>5 Normalized error for other cases</b>	<b>7</b>
<b>6 Comparison to transient climate run simulation at high latitude</b>	<b>9</b>
<b>7 Emulator products</b>	<b>10</b>

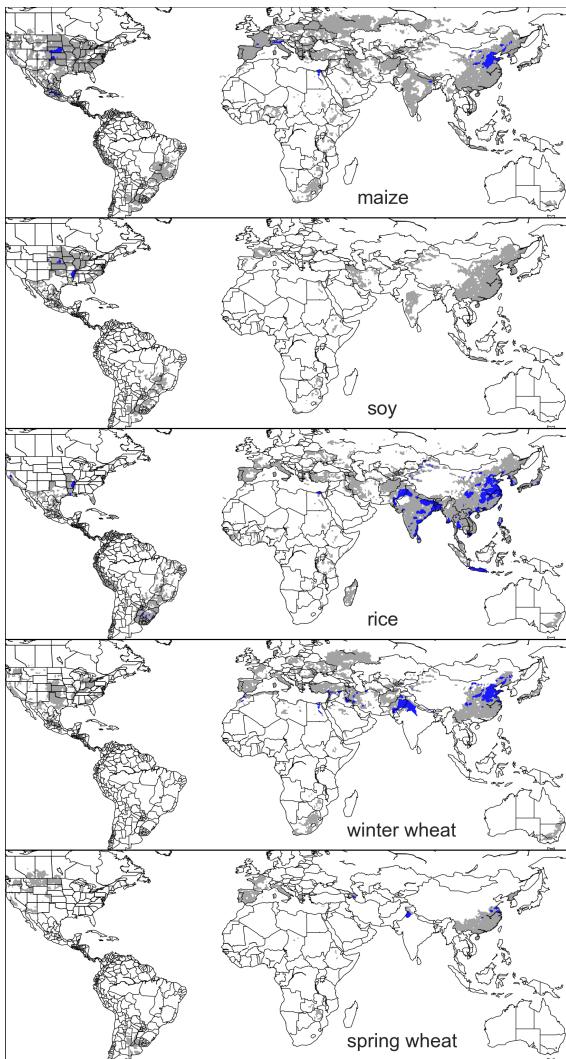
<b>8 Reduced specification (23-term) emulator examples</b>	<b>13</b>
<b>9 Yield Responses for other crops and models</b>	<b>18</b>
<b>10 Normalized Error for all models</b>	<b>20</b>
<b>11 Cross validation error for all models</b>	<b>24</b>

# 1 Experiment simulation sampling in variable space

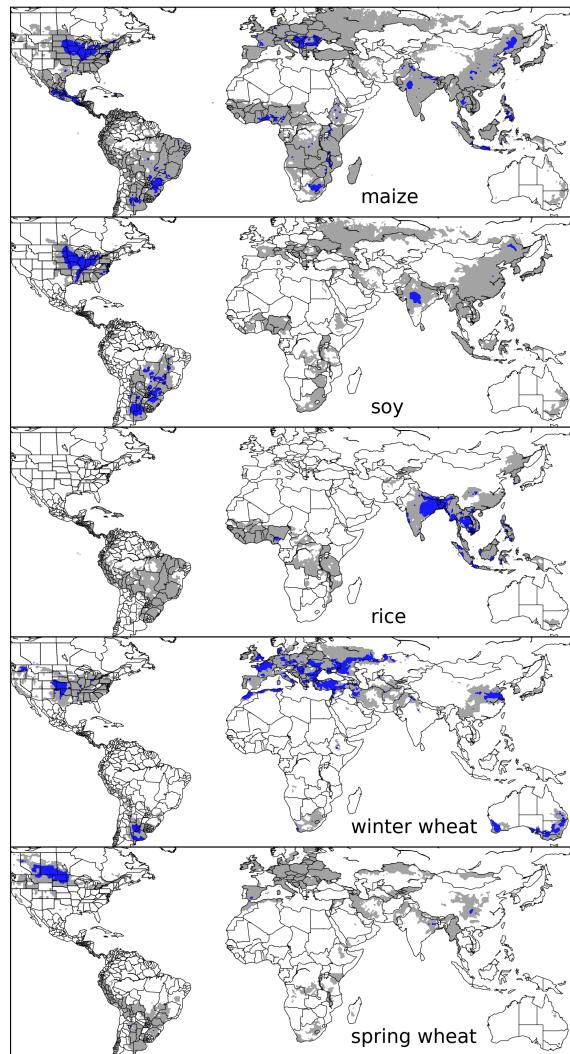


**Figure S1:** Heatmap illustrates number of model simulations provided for each of the scenarios in the variable space. The maximum number is 9, the number of models included in the emulator analysis (excluding three models not included in the emulator analysis). Normalized error calculations are run over scenarios with maximum number of models.

## 2 Cultivated areas

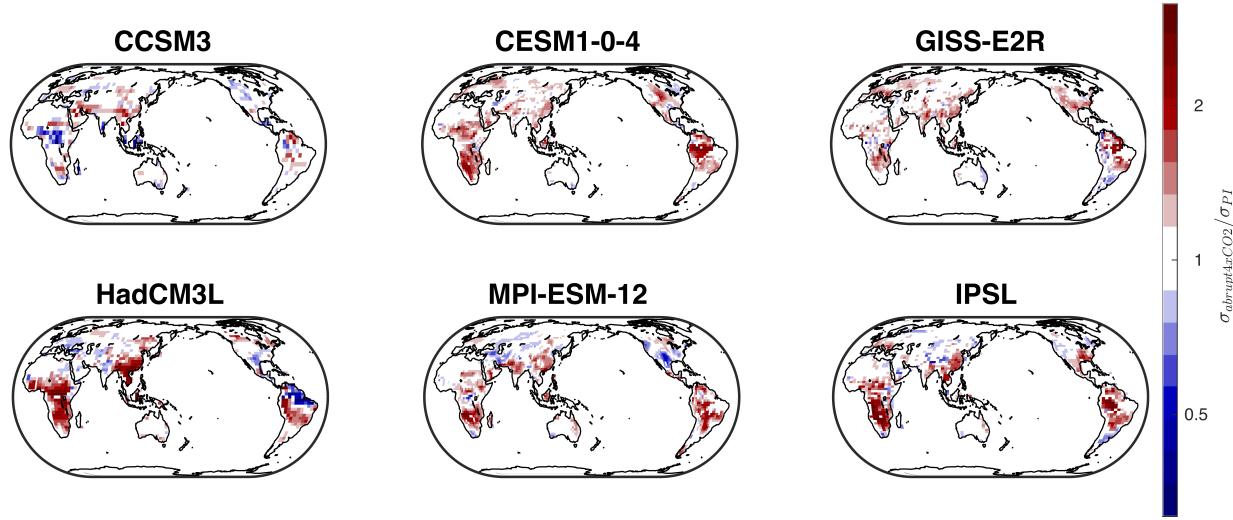


**Figure S2:** Presently cultivated area for irrigated crops in the real world. The blue contour area indicates grid-cells with more than 20,000 hectares of crop cultivated. The gray contour shows area with more than 10 hectares cultivated. Data from the MIRCA2000 data set for maize, rice, and soy. Winter and spring wheat areas are adapted from MIRCA2000 data and sorted by growing season



**Figure S3:** Presently cultivated area for rainfed crops in the real world. Conventions as in Figure S1.

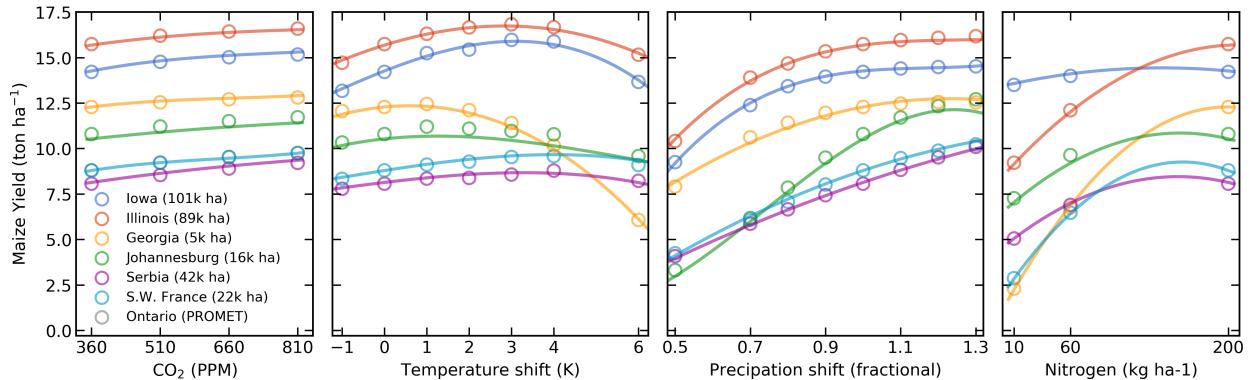
### 3 Variability changes in future climate projections



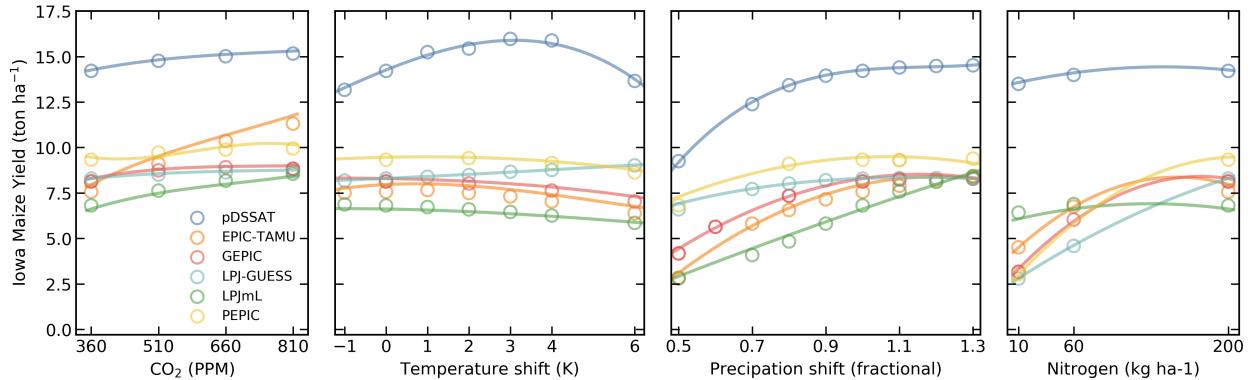
**Figure S4:** Change in hemispheric-summer temperature variability for selected models in the CMIP-5 archive for an abrupt 4xCO<sub>2</sub> forcing. Heatmap shows the ratio in standard deviation in temperature compared to preindustrial values, using 30-year detrended timeseries from JJA (northern hemisphere) and DJF (southern hemisphere). Climate model disagreement about changes in variability is substantial, though most project slight increases during summer. The HadCM3 model, used later in examples here, exhibits relatively strong variability changes, with strong increases in Central and South America, East Asia, and Africa.

## 4 Yield response for A1 (growing season adaptation) simulations

This section shows parallel figures to the main text for A1 simulations. Temperature responses are less for A1 simulation compared to A0 simulations because growing seasons are modified under warming via genetics in the A1 simulations.



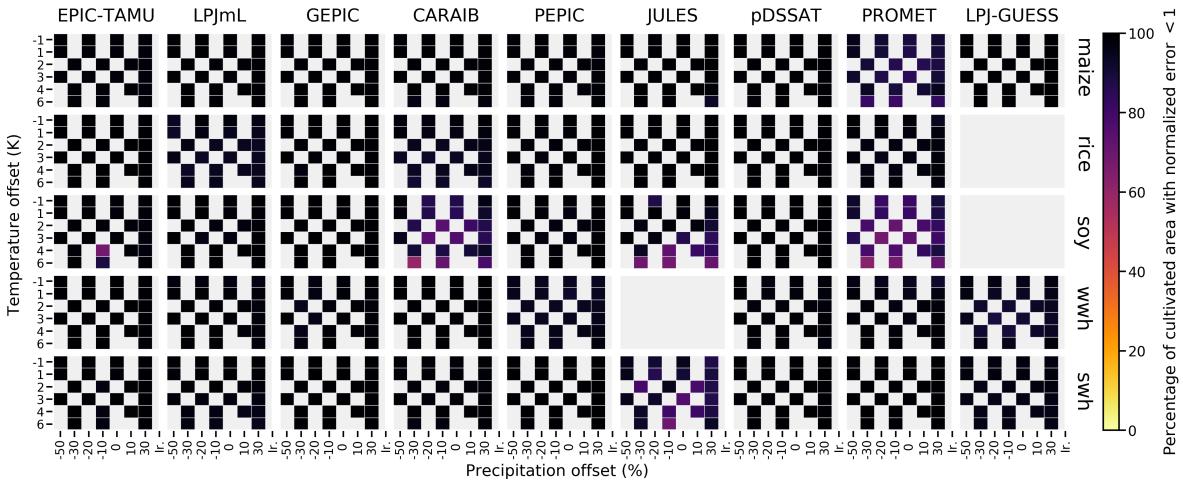
**Figure S5:** Illustration of spatial variations in yield response, which are successfully captured by the emulator for the A1 simulations. Panels show simulations (points) and emulations (lines) of rainfed maize in the pDSSAT model in six example locations selected to represent high-cultivation areas around the globe. Legend includes hectares cultivated in each selected grid cell. Each panel shows variation along a single variable, with others held at baseline values.



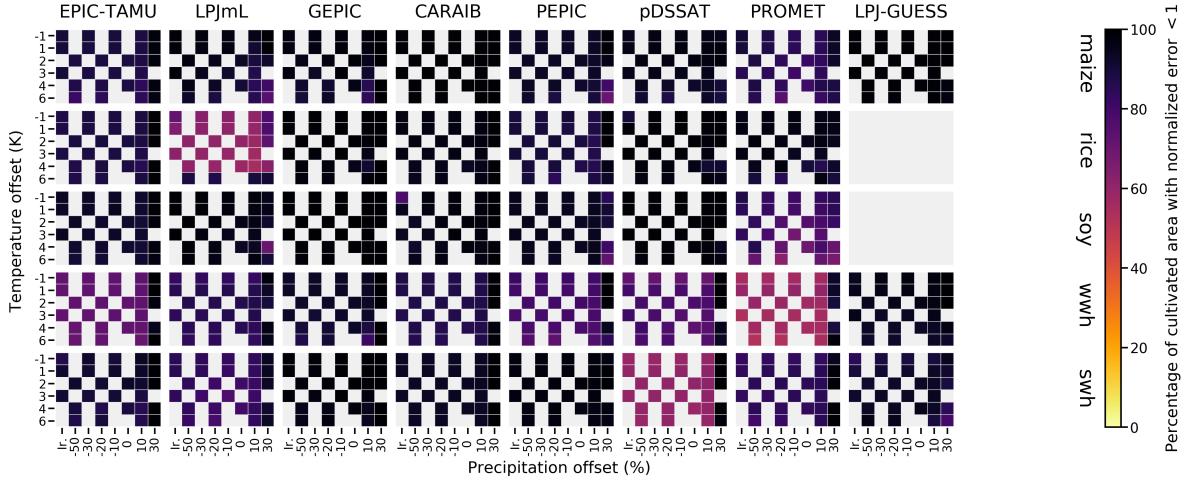
**Figure S6:** Illustration of variations in yield response across models for A1 simulations, again successfully captured by the emulator. Panels show simulations and emulations from six representative GCM models for rainfed maize in the same Iowa grid cell shown above, with the same plot conventions. Three models (PROMET, JULES, and CARAIB) that do not simulate the nitrogen dimension are omitted for clarity.

## 5 Normalized error for other cases

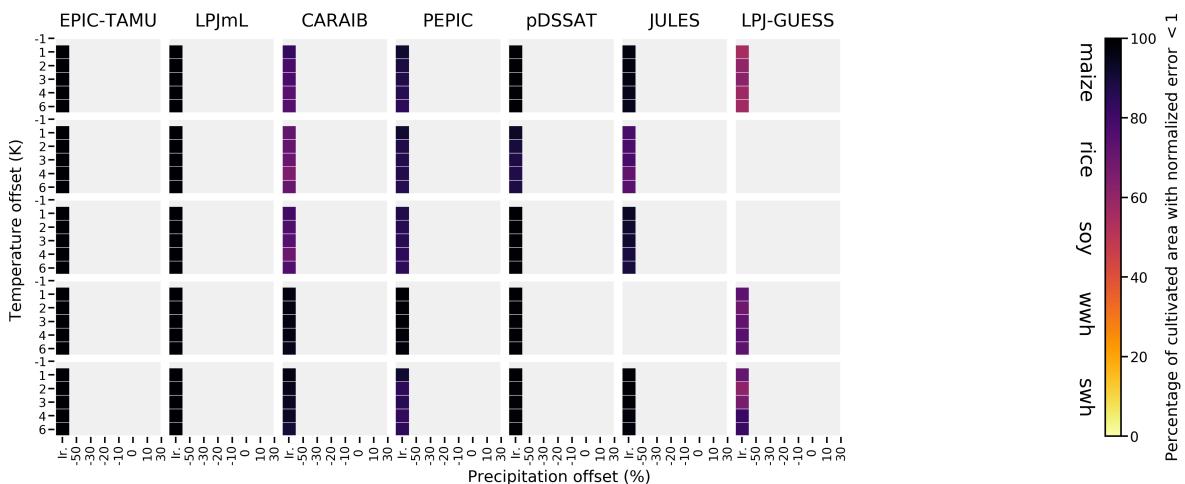
In manuscript Figure 7 we show normalized error for the A0 emulators over all rainfed crops, models, and T and W values for baseline CO<sub>2</sub> and nitrogen levels (360 ppm and 200 kg ha<sup>-1</sup>). Here we show normalized error for the A0 emulators of rainfed crops at higher CO<sub>2</sub> (Figure S6), for the A1 emulators of rainfed crops at baseline values (Figure S7), and for the A0 emulators of irrigated crops at baseline values (Figure S8). Results are generally similar, except normalized errors at higher CO<sub>2</sub> are generally lower because model disagreement is larger, and some model emulators for irrigation water demand are under-performing (LPJ-GUESS and CARAIB for some crops). A1 errors are similar to A0 errors across crops and models with a few exceptions: LPJmL rice, pDSSAT spring wheat, and PROMET winter wheat.



**Figure S7:** The fraction of currently cultivated hectares with normalized emulation error less than 1 for the CO<sub>2</sub>=810 ppm and 200 kg N ha<sup>-1</sup> yr<sup>-1</sup> case for the temperature and precipitation perturbations scenarios provided by all 9 models included in the emulator analysis. Figure convention is as in main text Figure 7. The yield response is generally easy to emulate over currently cultivated areas (black regions).



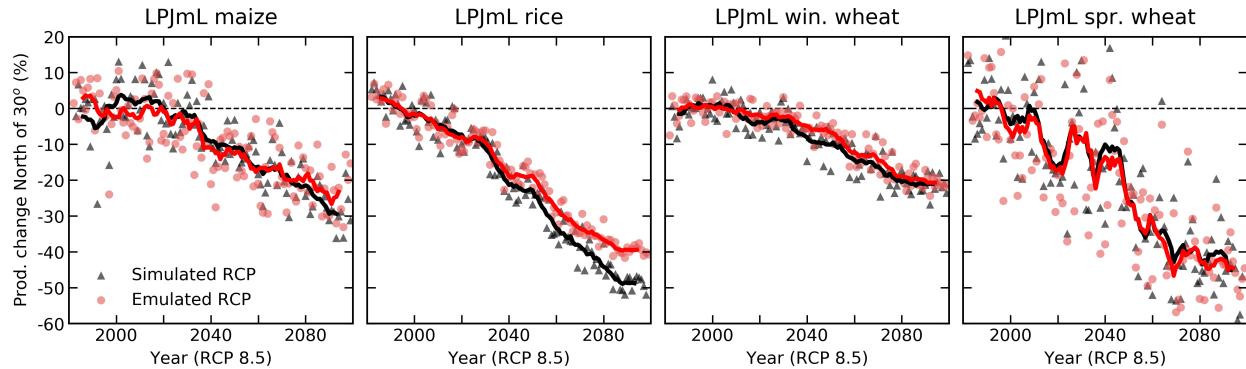
**Figure S8:** The fraction of currently cultivated hectares with normalized emulation error less than 1 for A1 yield emulation for  $\text{CO}_2=310 \text{ ppm}$  and  $200 \text{ kg N ha}^{-1} \text{ yr}^{-1}$  case. Figure convention is the same as main text Figure 7.



**Figure S9:** The fraction of currently cultivated hectares with normalized emulation error less than 1 for irrigated water demand emulation for  $\text{CO}_2=310 \text{ ppm}$  and  $200 \text{ kg N ha}^{-1} \text{ yr}^{-1}$  case. Figure convention is the same as main text Figure 7.

## 6 Comparison to transient climate run simulation at high latitude

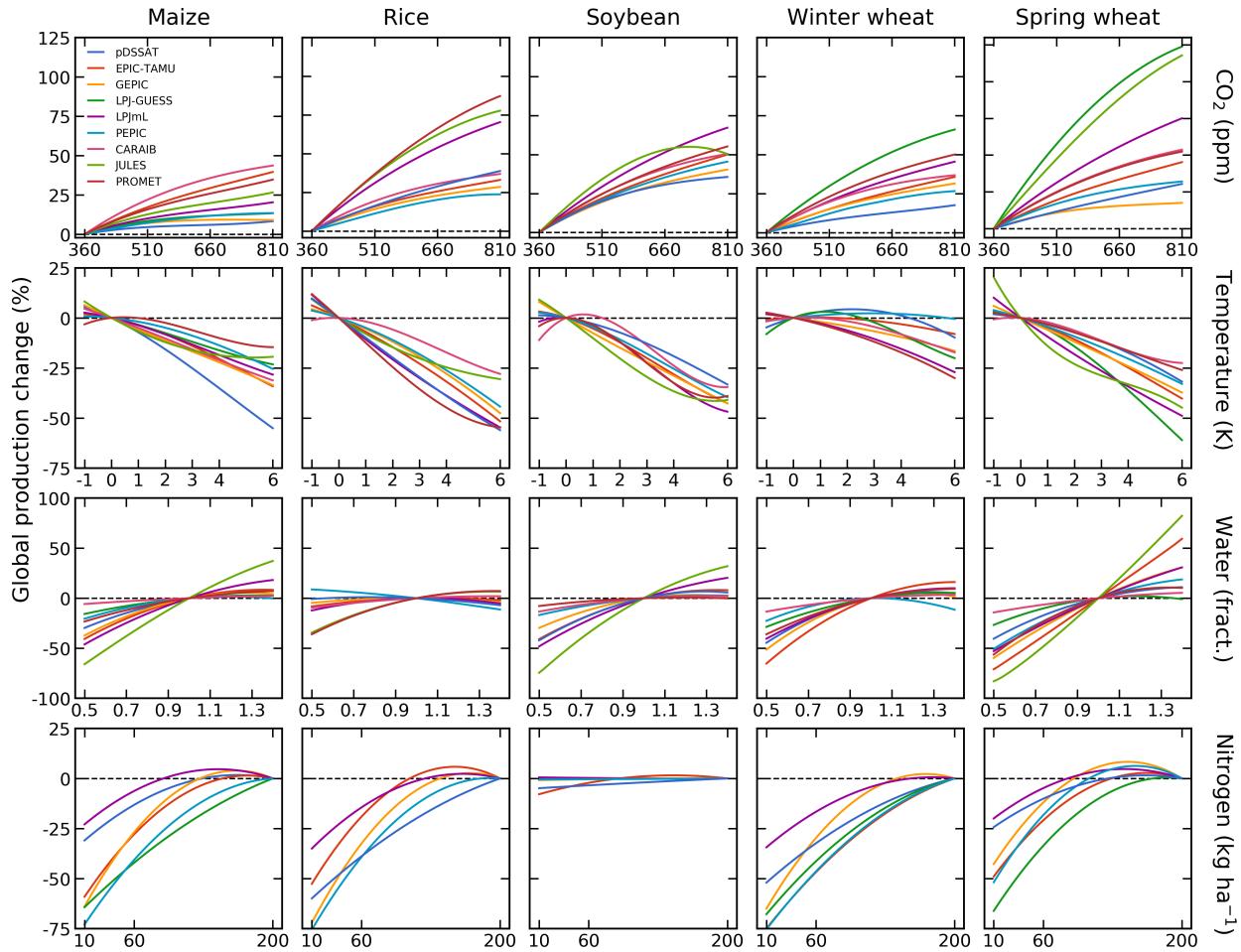
We test the emulator against crop model simulations driven by a transient climate run (which contains some variability changes) at high latitude to gage the impact in variability changes that are not captured by the emulator. High latitudes are isolated here in contrast to Figure 9 in the main text because changes in variability are thought to matter most in this region. Changes in variability are matter little for all crops except rice north of 30N latitude. Very little rice is grown in the higher latitudes (20% north of 30N and 1% north of 45N).



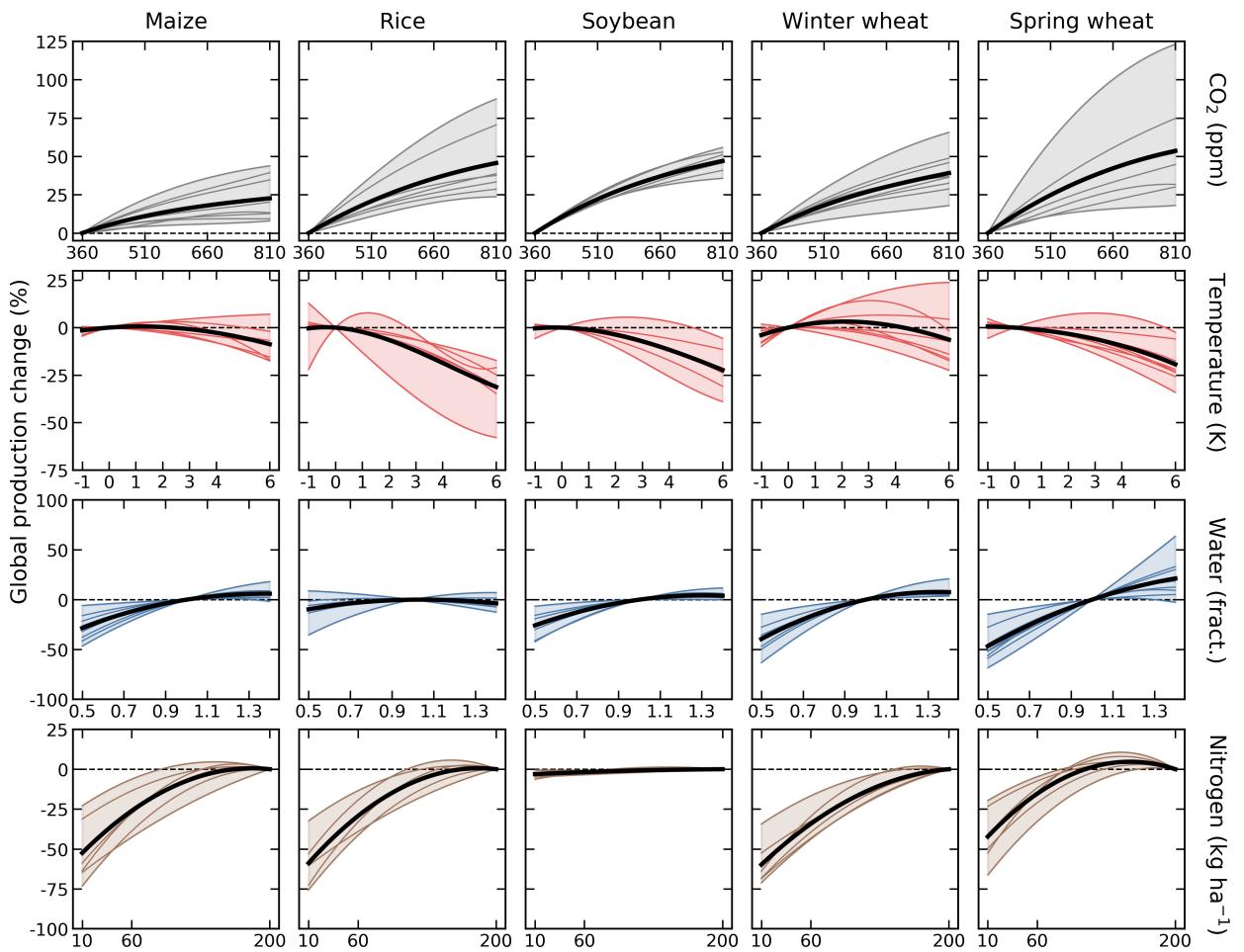
**Figure S10:** Illustration of the ability of the emulator to capture an RCP simulation, again for the HadCM model, for crops north of 30N latitude. Figure convention is the same as main text Figure 9.

## 7 Emulator products

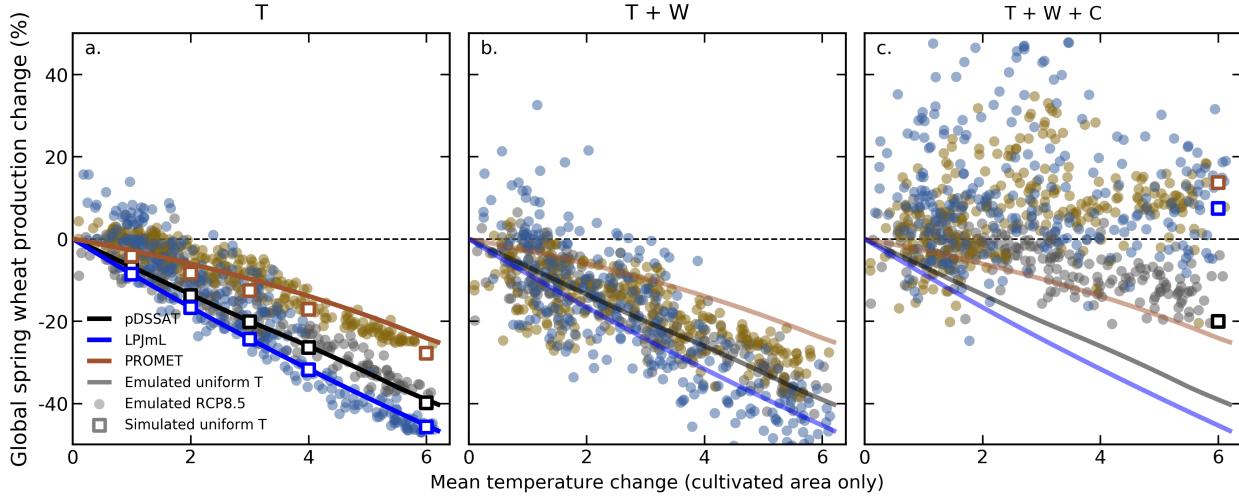
In this section we present some additional emulator products to illustrate potential emulator applications.



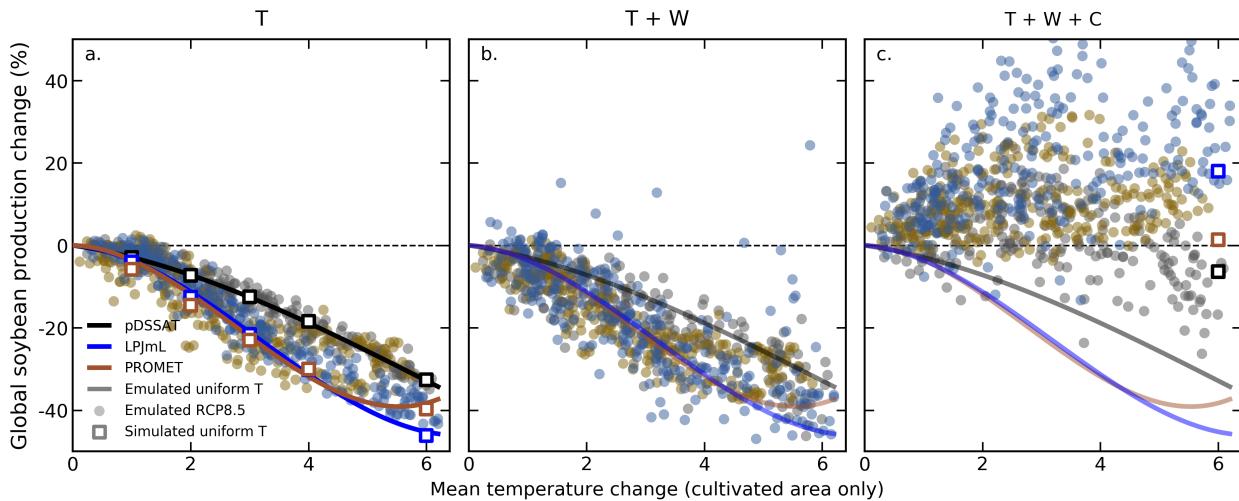
**Figure S11:** Same as main text Figure 10 except now each model is shown in color.



**Figure S12:** Emulated damage functions for rainfed crops for A1 (growing season adaptation) simulations. Note: JULES does not provide A1 simulations. Same convention as main text Figure 10. Temperature responses are generally flatter than for A0 simulations.



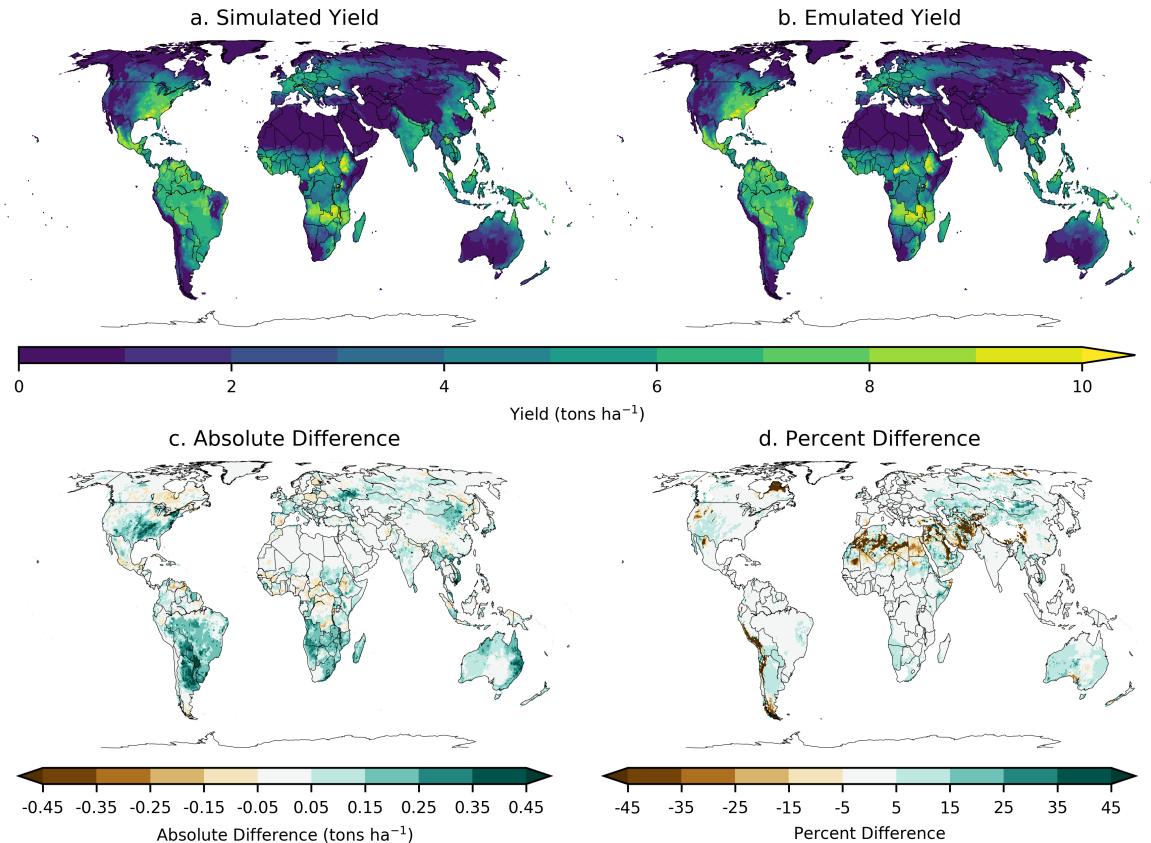
**Figure S13:** Illustration of the factors affecting yields in more realistic climate scenarios for rainfed and irrigated (current mix) spring wheat. Conventions as in main text Figure 11. Large emulator errors in PROMET spring wheat temperature response (panel **a**, compare open squares to line) are driven by Southern China, where discontinuities in yield responses make emulation problematic. (See Figure in Supplemental Material Section S10).



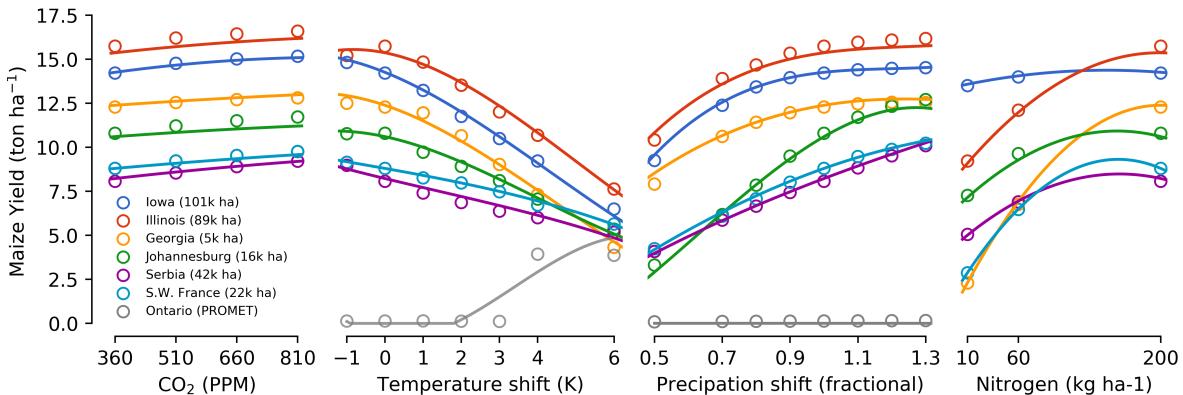
**Figure S14:** Illustration of the factors affecting yields in more realistic climate scenarios for rainfed and irrigated (current mix) soy. Conventions as in main text Figure 11. The split in PROMET soybean temperature response (panel **a**, note distinct groups of points) results from the model's sensitivity to differences in spatial patterns of temperature change across climate models.

## 8 Reduced specification (23-term) emulator examples

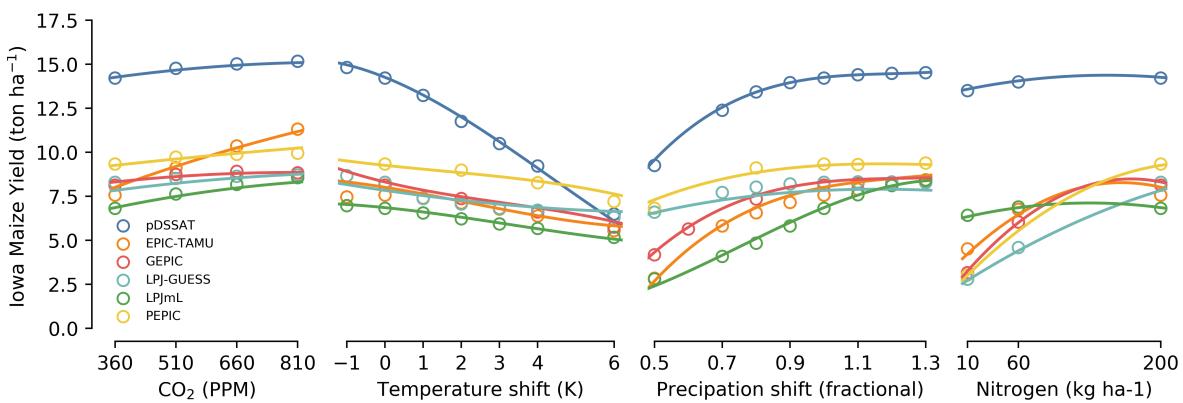
In this section we present parallel figures for the reduced-form emulator (23-term) for reference. Issues with the reduced-form model are most prominent in PROMET for rice and soy, and JULES soy and spring wheat. Some potential causes behind this difference in ease of emulation are as follows. First, we do not emulate PROMET or JULES in the nitrogen dimension (CARAIB is the other model that does not do N). Both JULES and PROMET models are land system process models, originally developed with a broad focus, and have relatively recently (2015) been adapted for managed vegetation (i.e. agriculture). PROMET is the most sensitive model of all the models for rice in C, T, and W. PROMET is the quantitatively lowest performing model for soybeans when compared to the historical FAO data for the top 10 producing countries. JULES is the most sensitive model of all the models for soybeans in C, T, and W. For spring wheat, JULES is a high outlier in C, the most sensitive model in W and T, and shows an extra inflection point in the global temperature response not seen in any of the other models. CARAIB on the other hand (the other model that does not do nitrogen), was originally developed as a vegetation model in the early 90's and has a longer history of agricultural focus. It is the least sensitive model in T for rice and spring wheat, and is nearest to the ensemble mean in maize and winter wheat. In W, it is the least sensitive model for maize, and the wheats, the second least sensitive model for soybeans, and nearest to the ensemble mean for rice. For C, it is the nearest to the ensemble mean for the wheats and soy and unremarkable for rice. However, CARAIB is the most sensitive model for maize in the C dimension (though this is relatively less sensitive than most other responses for other models the other crops).



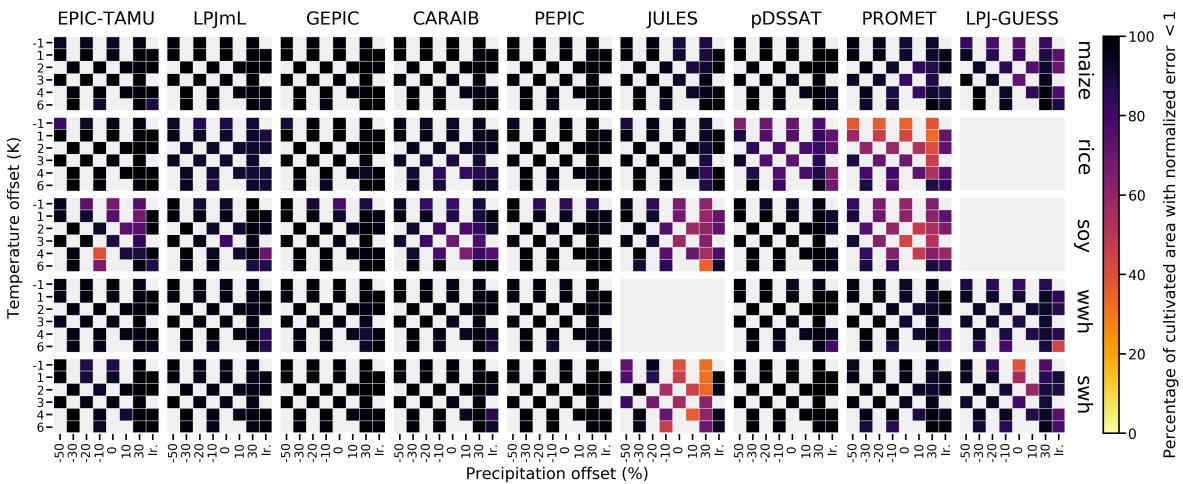
**Figure S15:** Same convention as Figure 4 in the main text except now with the reduced (23-term) emulator specification.



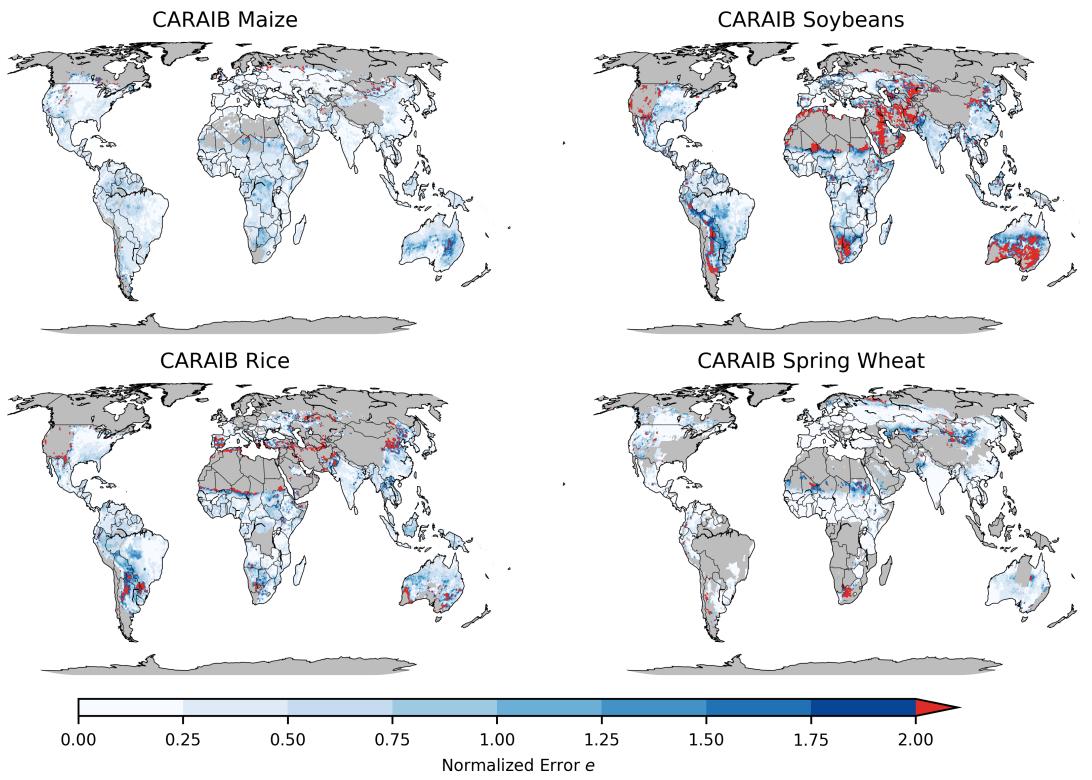
**Figure S16:** Same convention as Figure 5 in the main text except now with the reduced (23-term) emulator specification.



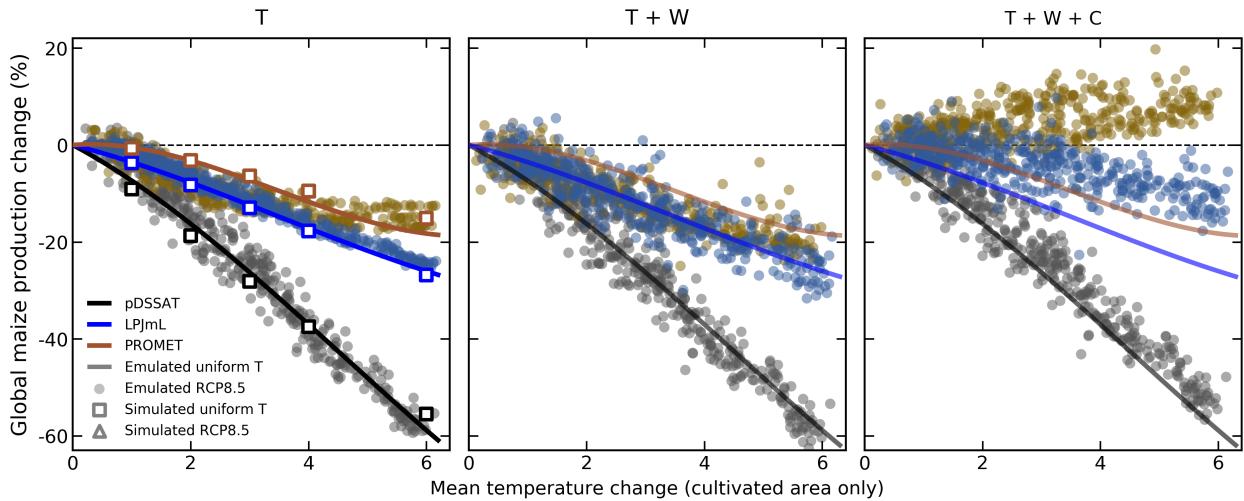
**Figure S17:** Same convention as Figure 6 in the main text except now with the reduced (23-term) emulator specification.



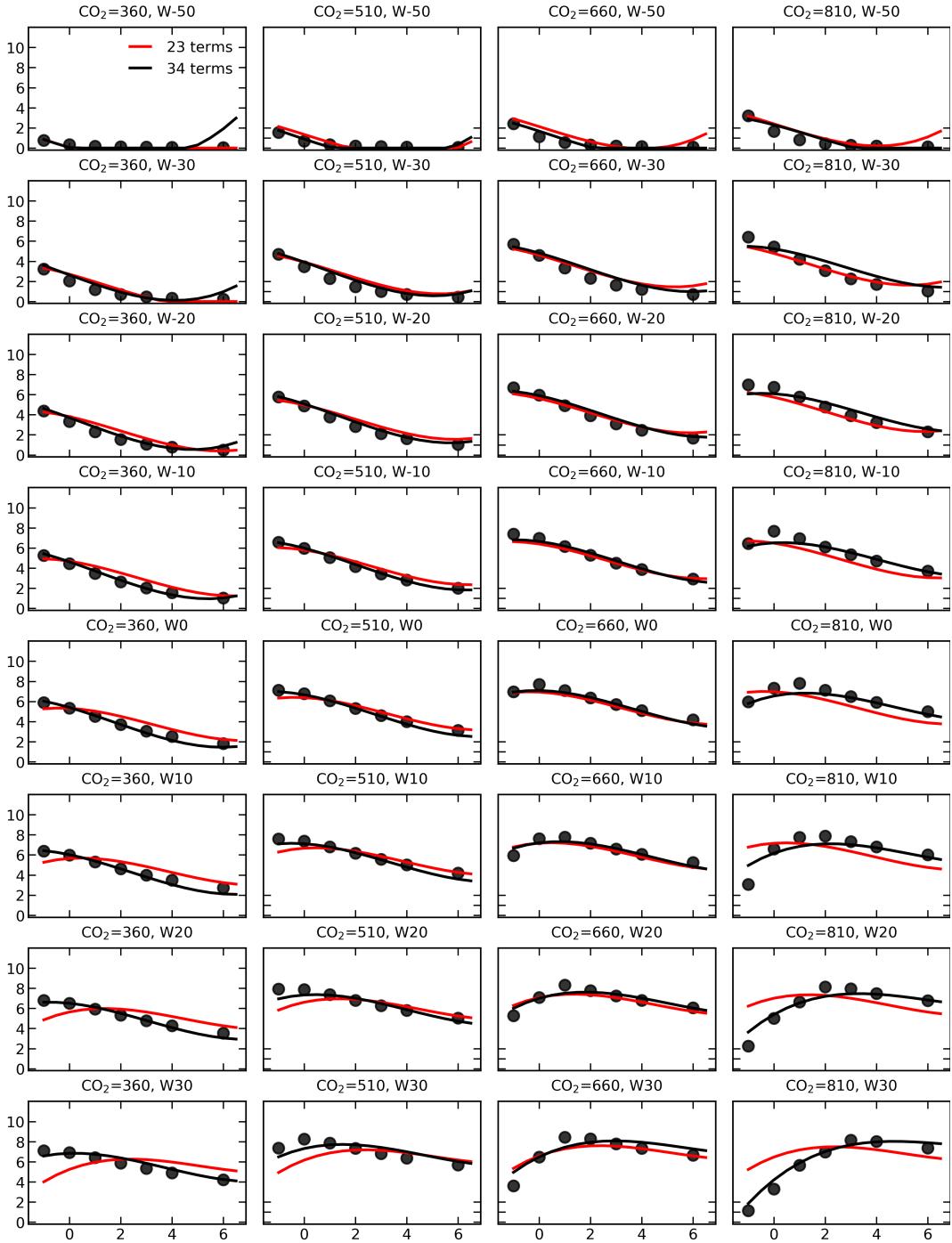
**Figure S18:** Same convention as Figure 7 in the main text except now with the reduced (23-term) emulator specification.



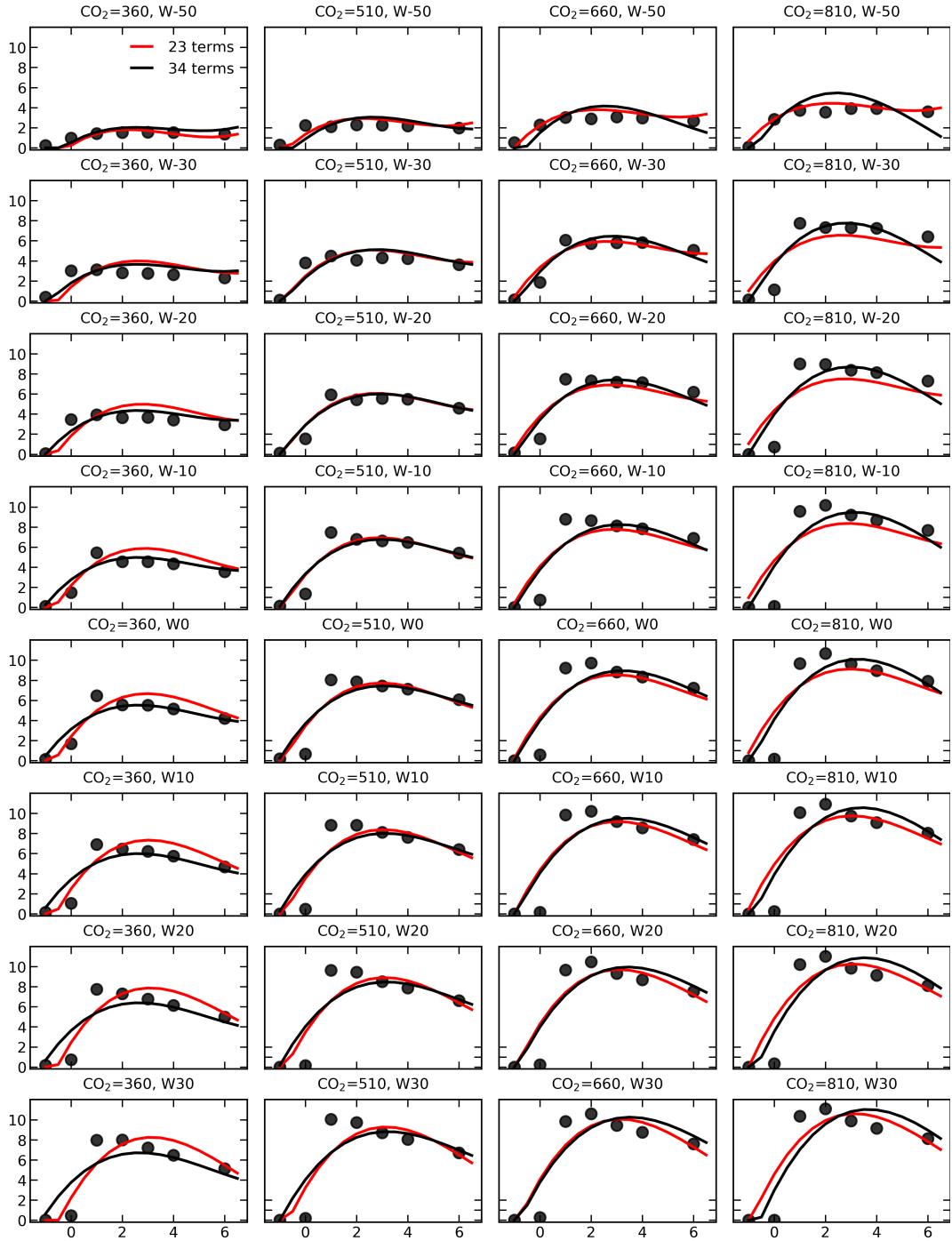
**Figure S19:** Same convention as Figure 8 in the main text except now with the reduced (23-term) emulator specification.



**Figure S20:** Same convention as Figure 11 in the main text except now with the reduced (23-term) emulator specification. The C response here is different than the full-form emulator because it does not contain many of the higher order C ( $C^3, C^2 * T \dots$ ) interaction terms. This results in substantial differences for PROMET.



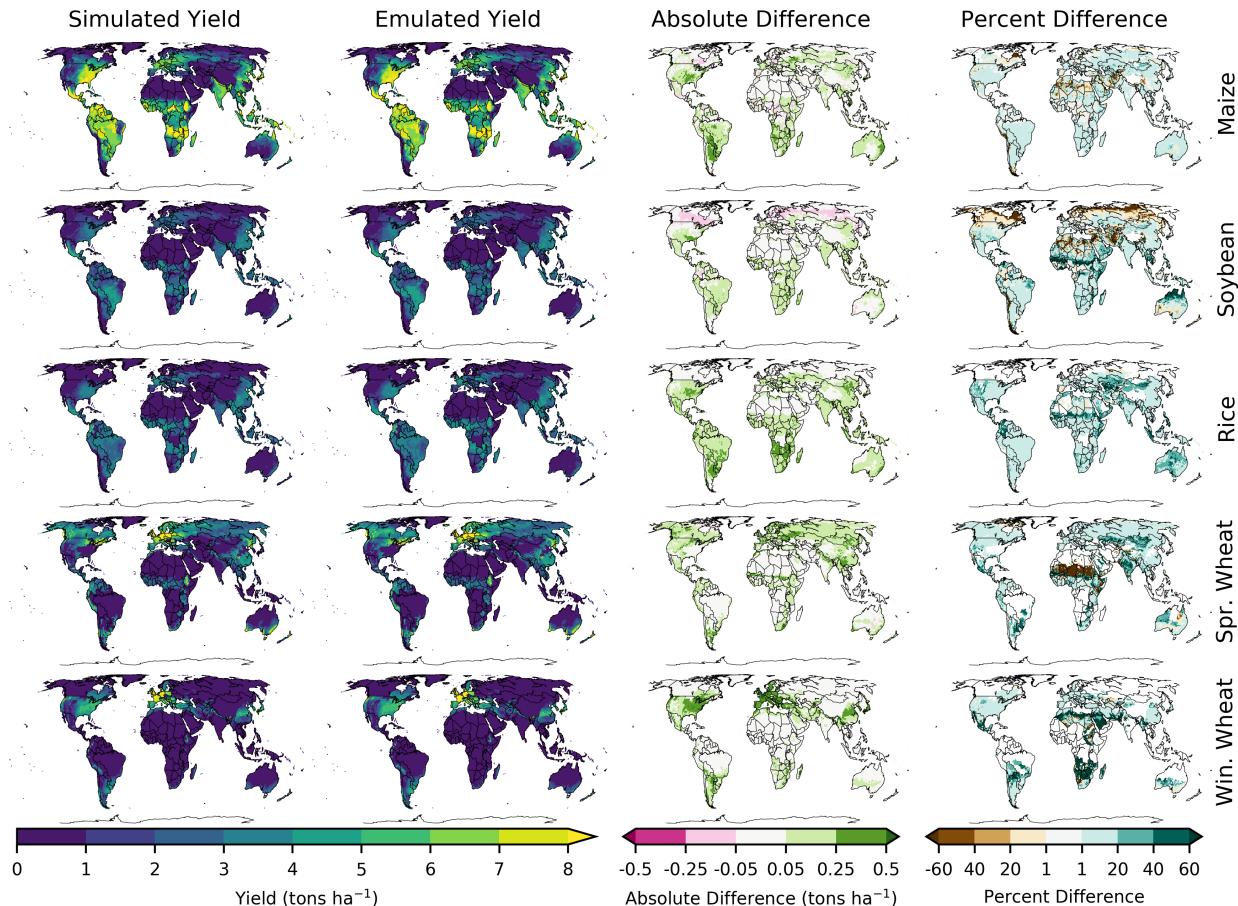
**Figure S21:** Example of emulator failure. Simulated and emulated values for JULES soybean in Southern Germany. RMSE = 41% of baseline yield for the reduced form (23-term) emulator. The downturn in yields as C and W increase can only be captured by the higher order C interaction terms.



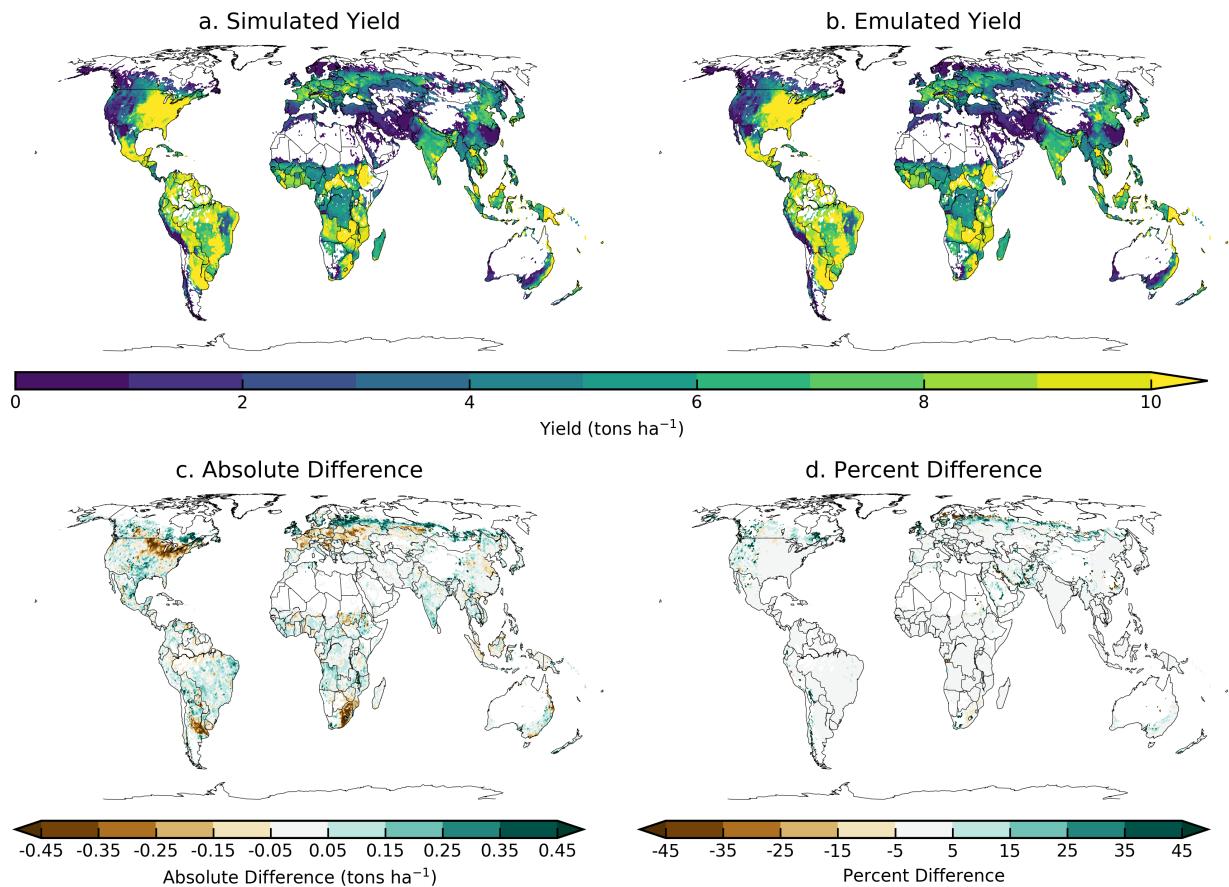
**Figure S22:** Example of emulator failure. Simulated and emulated values for PROMET rice in Arunachal Pradesh. RMSE = 132% of baseline yield for red (reduced fit). The step change in the yields around 0 K at higher water specifications cannot be captured by any third order polynomial. Both emulator specifications fail in this example.

## 9 Yield Responses for other crops and models

We present some additional yield response spatial patterns for other crops for the LPJmL model and for pDSSAT for maize for reference.



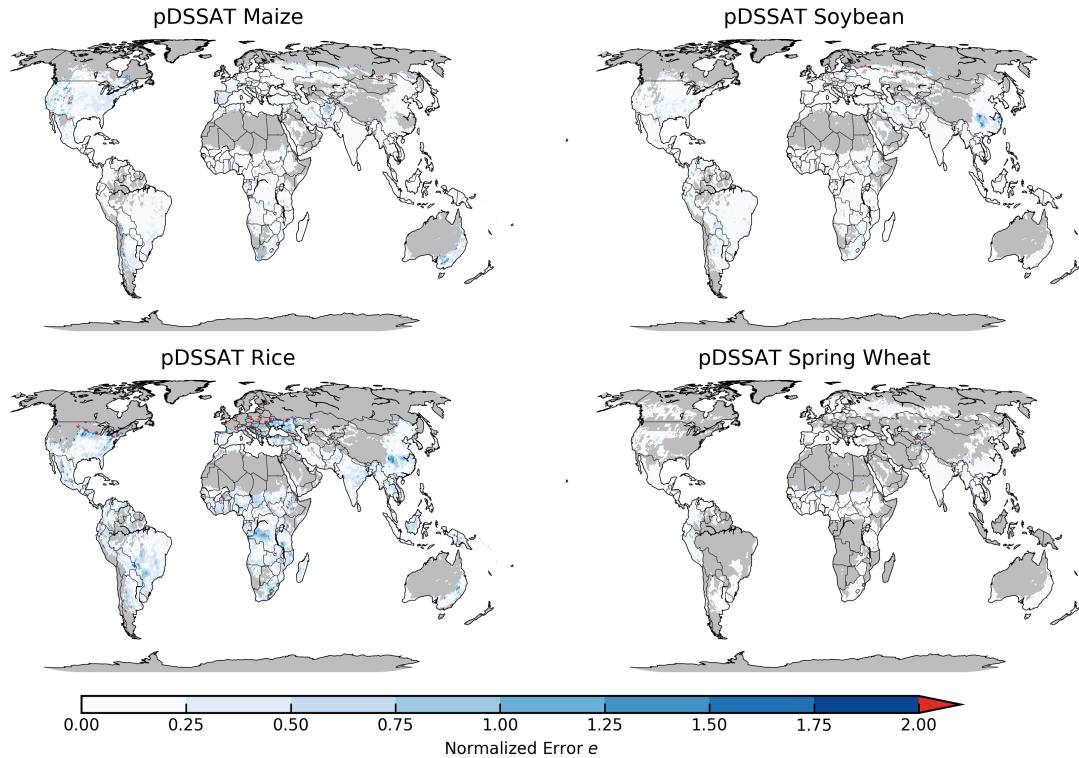
**Figure S23:** Spatial yield response and emulator error for LPJmL for all crops included in the stud. Same convention as Figure 4 in the main text.



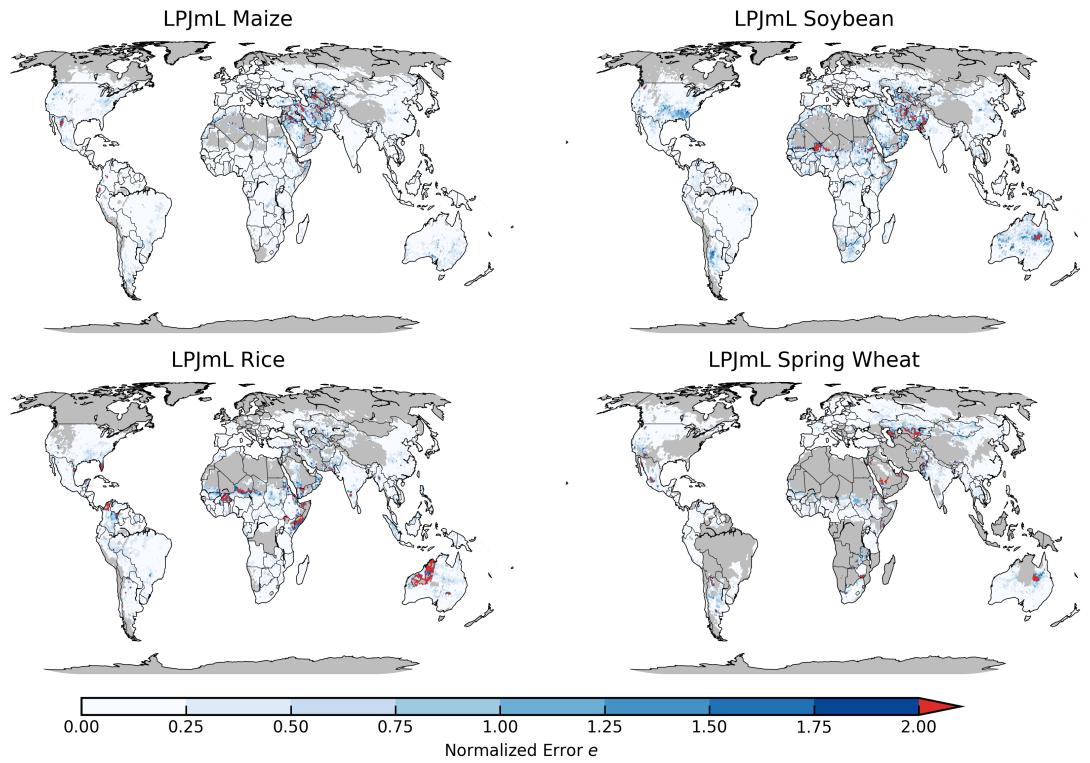
**Figure S24:** Spatial yield response and emulator error for pDSSAT for maize. Same convention as Figure 4 in the main text.

## 10 Normalized Error for all models

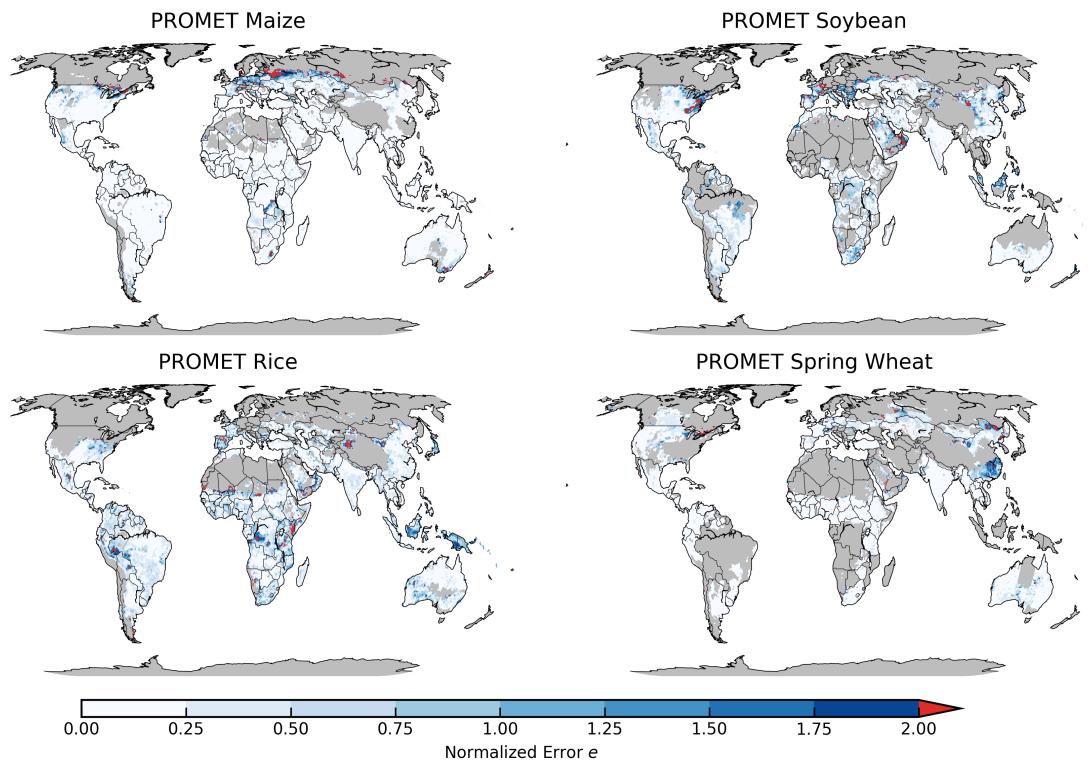
In this section we present maps of normalized error  $e$  with the same convention as Figure 8 in the main text for all models for completeness.



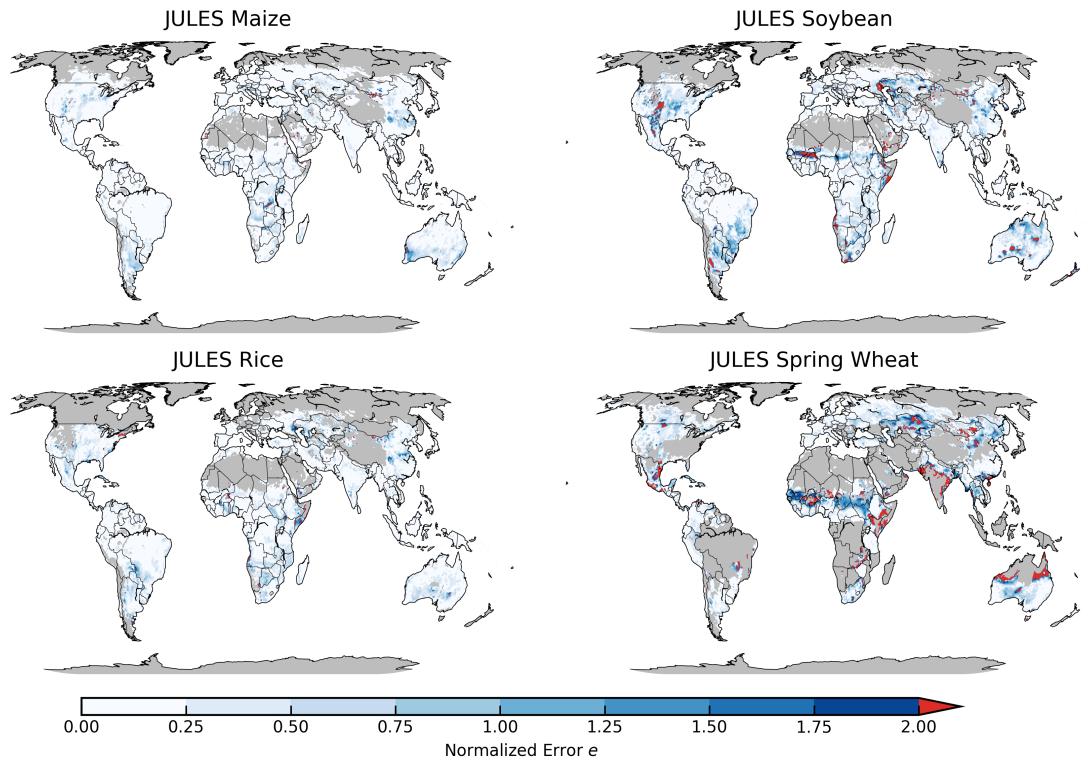
**Figure S25:** Normalized error  $e$  for pDSSAT. Same convention as main text Figure 8.



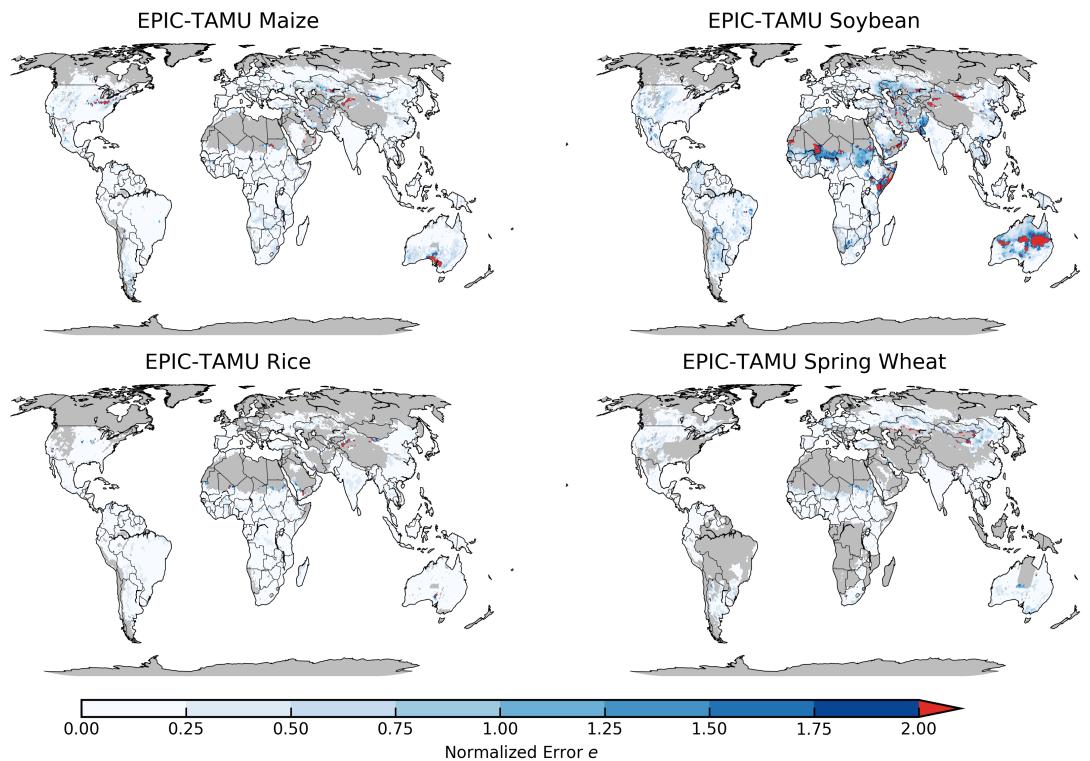
**Figure S26:** Normalized error  $e$  for LPJmL. Same convention as main text Figure 8.



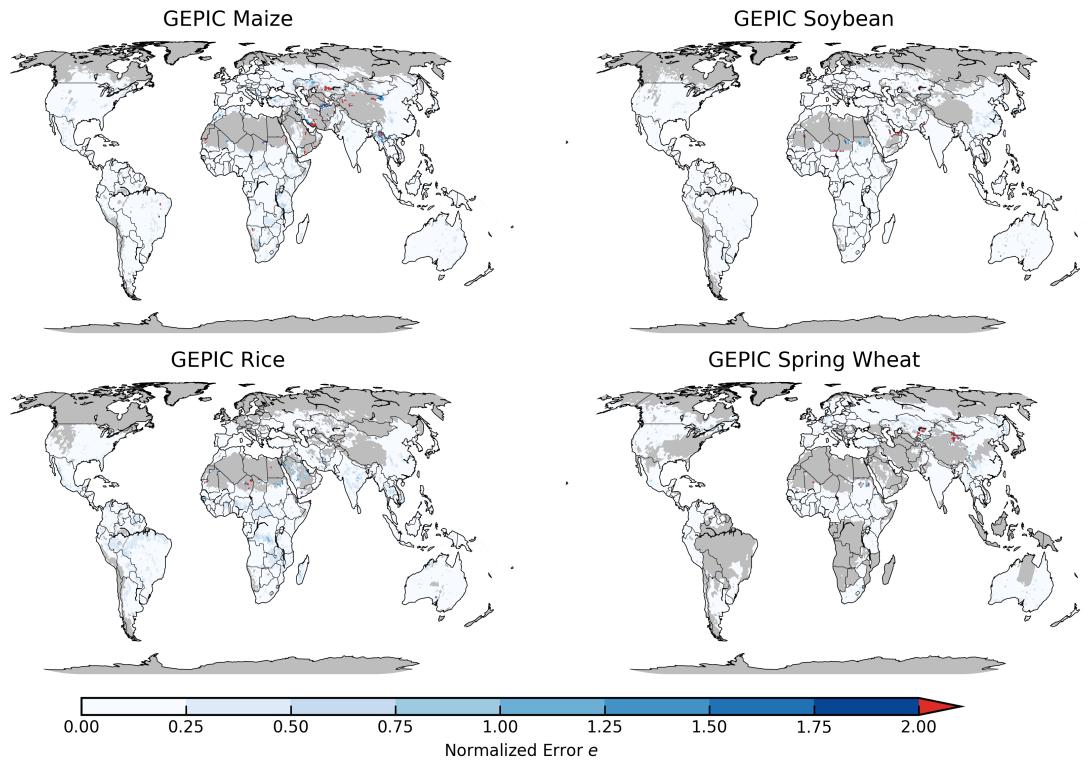
**Figure S27:** Normalized error  $e$  for PROMET. Same convention as main text Figure 8.



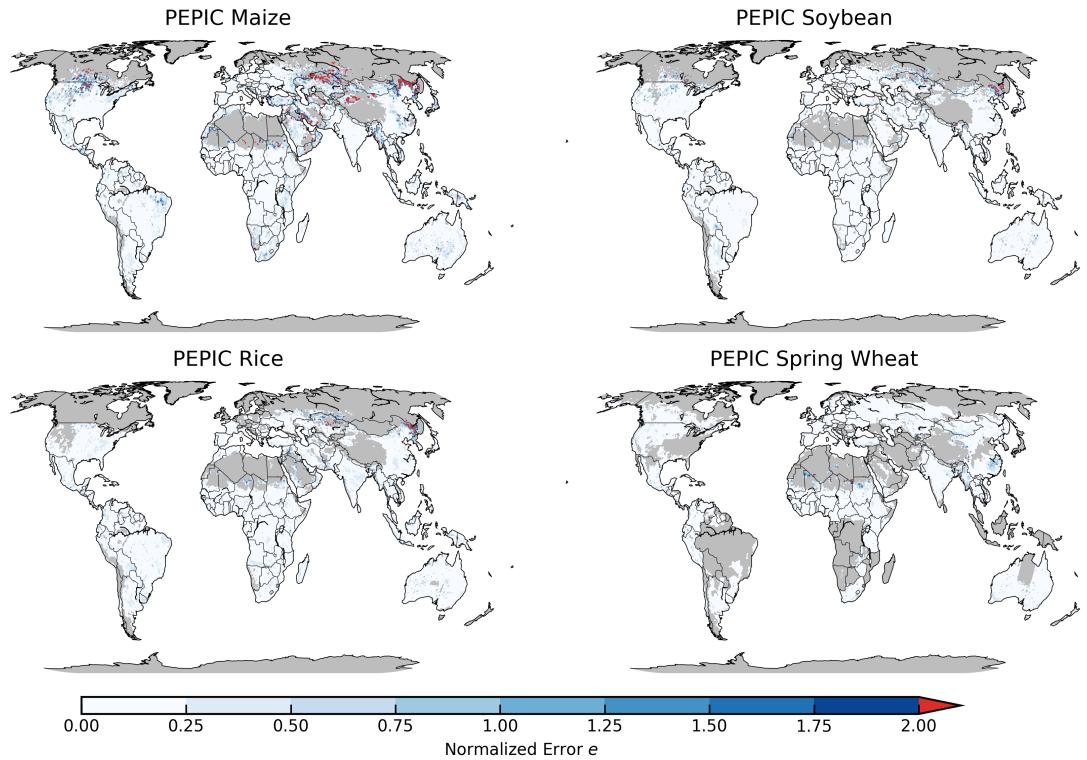
**Figure S28:** Normalized error  $e$  for JULES. Same convention as main text Figure 8.



**Figure S29:** Normalized error  $e$  for EPIC-TAMU. Same convention as main text Figure 8.



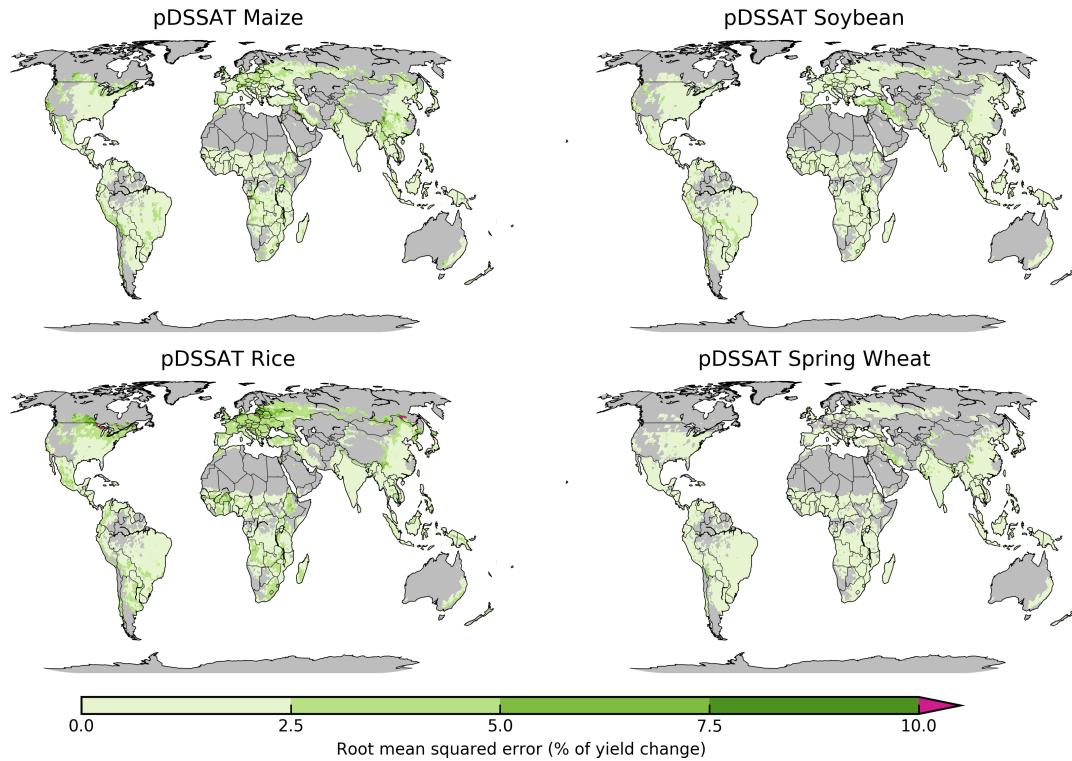
**Figure S30:** Normalized error  $e$  for GEPIC. Same convention as main text Figure 8.



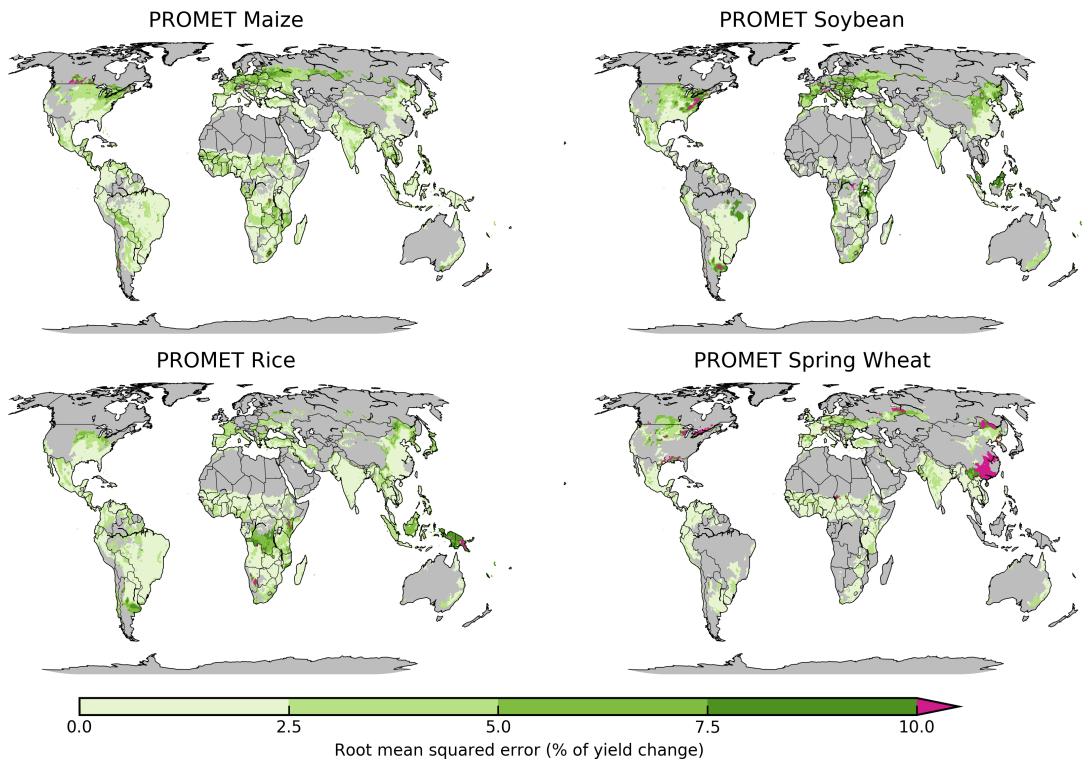
**Figure S31:** Normalized error  $e$  for PEPIC. Same convention as main text Figure 8.

## 11 Cross validation error for all models

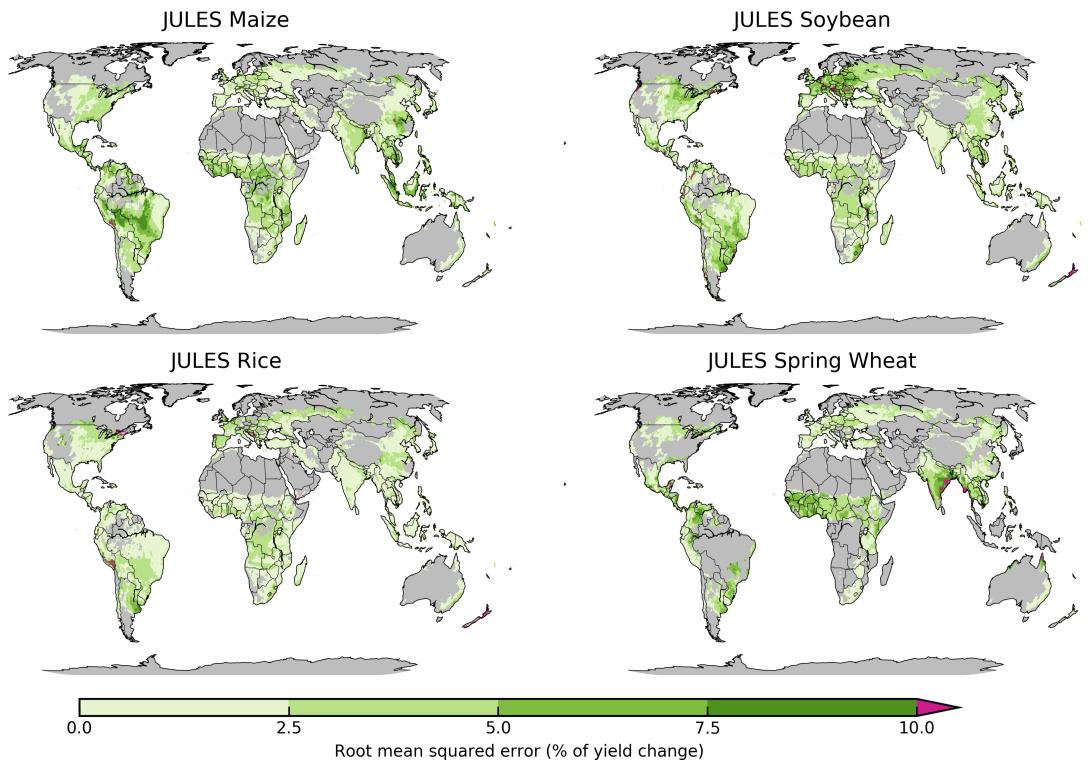
In this section we present maps of cross validation error (values found in main text Table 3 are aggregated up from the grid cell level). Errors are generally low as a percentage of yield change in each grid cell. Errors can be above 10% of yield change for the out-of-sample test for



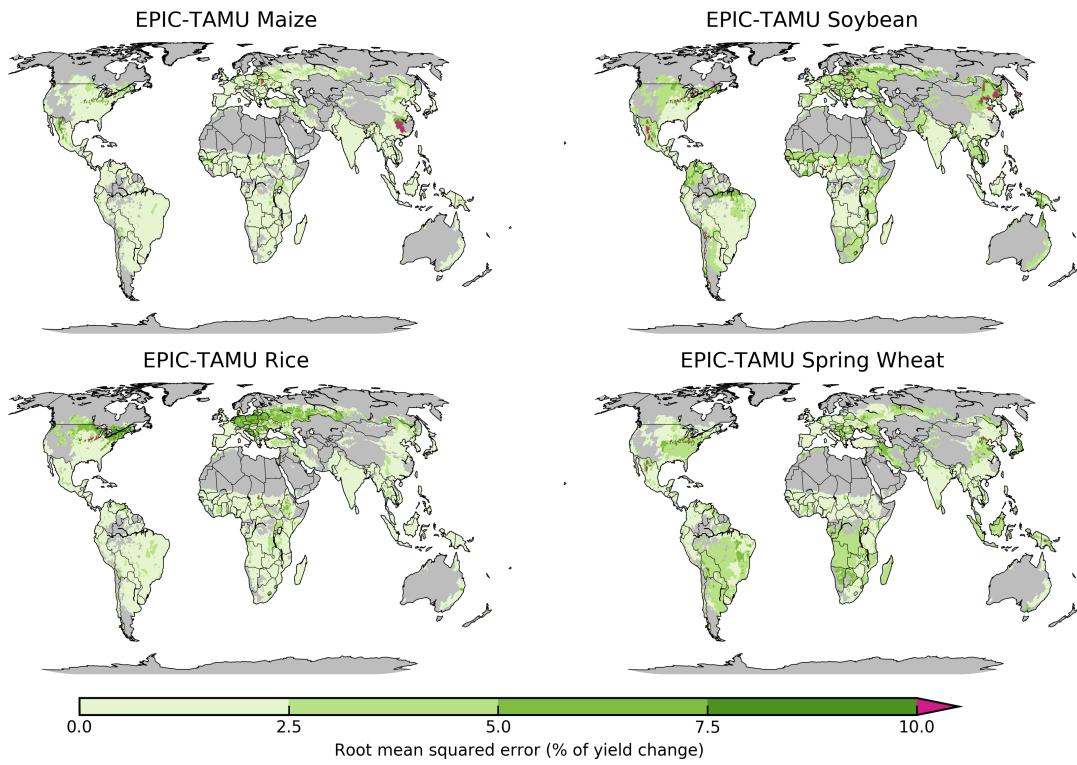
**Figure S32:** Root mean squared error for three-fold cross validation for the pDSSAT model for rainfed crops. Values shown as a percentage of yield change in each grid cell.



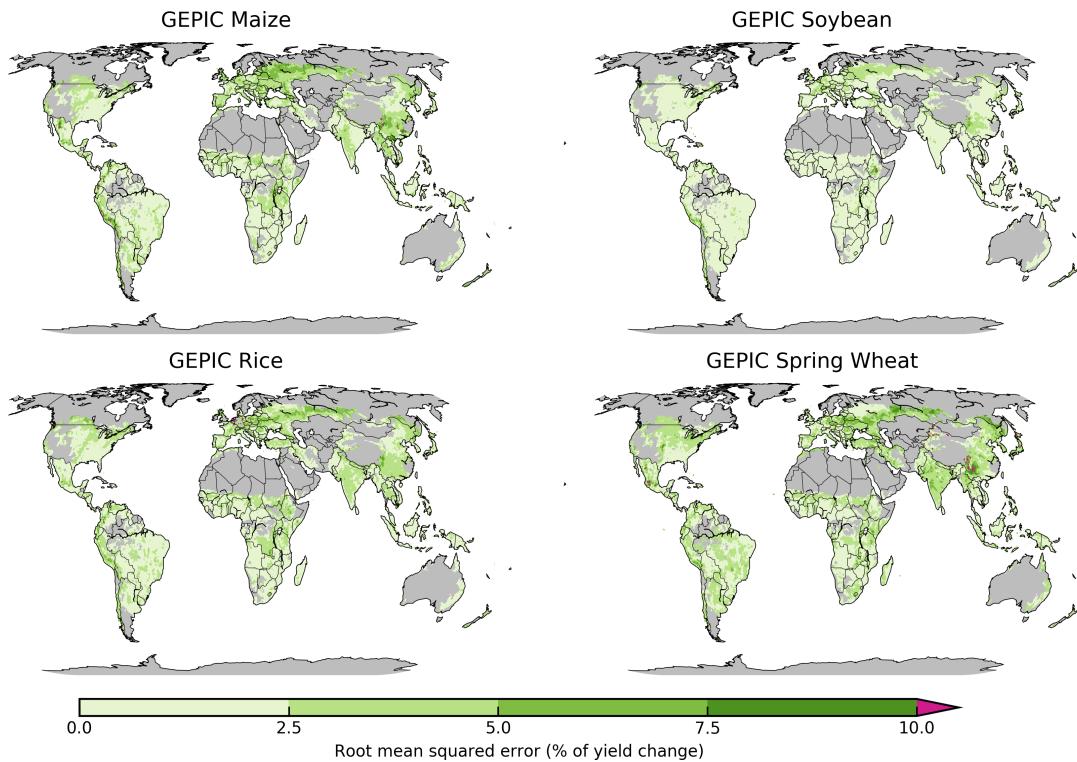
**Figure S33:** Map of root mean squared error for three fold cross validation process for the PROMET model for rainfed crops. Values shown as a percentage of yield change in each grid cell.



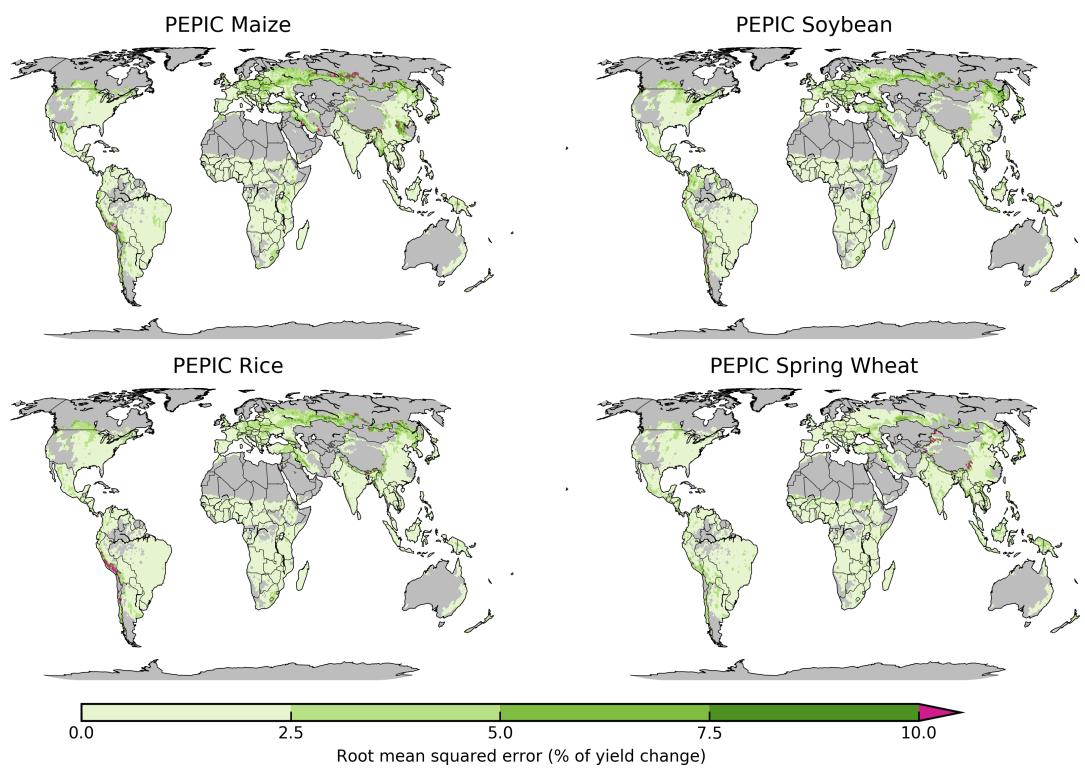
**Figure S34:** Map of root mean squared error for three fold cross validation process for the JULES model for rainfed crops. Values shown as a percentage of yield change in each grid cell.



**Figure S35:** Map of root mean squared error for three fold cross validation process for the EPIC-TAMU model for rainfed crops. Values shown as a percentage of yield change in each grid cell.



**Figure S36:** Map of root mean squared error for three fold cross validation process for the GEPIC model for rainfed crops. Values shown as a percentage of yield change in each grid cell.



**Figure S37:** Map of root mean squared error for three fold cross validation process for the PEPIC model for rainfed crops. Values shown as a percentage of yield change in each grid cell.