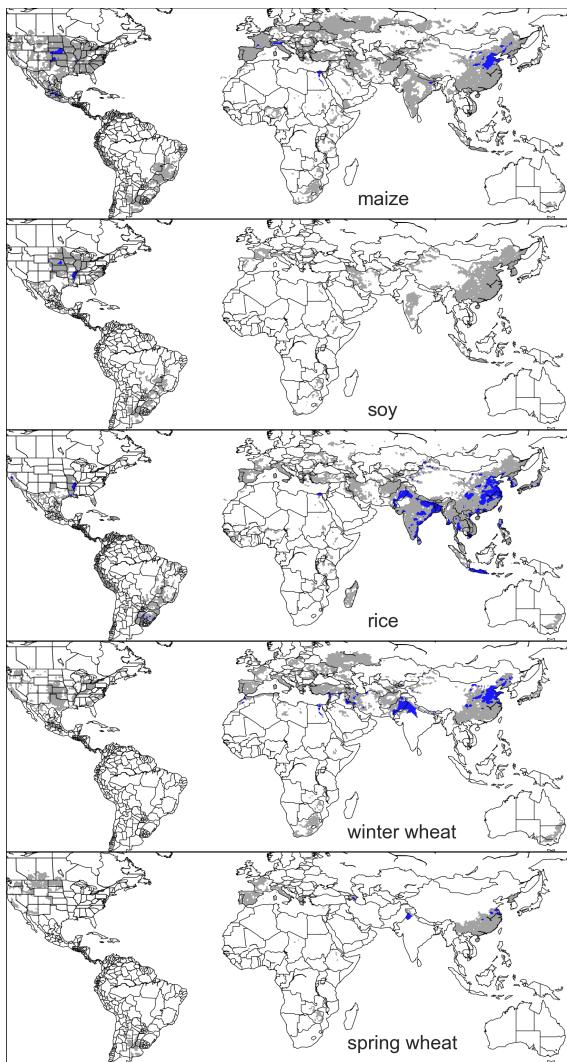


**Supplemental Material**  
**The GGCMI Phase II experiment: simulating and emulating  
global crop yield responses to changes in carbon dioxide,  
temperature, water, and nitrogen levels**

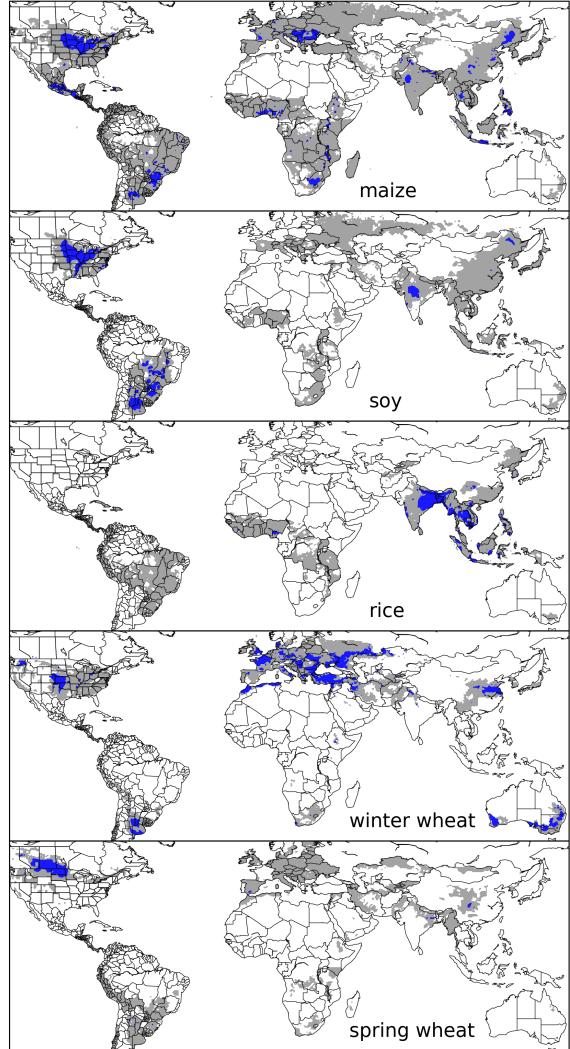
James Franke<sup>1,2</sup>, Joshua Elliott<sup>2,3</sup>, Christoph Müller<sup>4</sup>, Alexander Ruane<sup>5</sup>, Abigail Snyder<sup>6</sup>,  
Jonas Jägermeyr<sup>3,2,4,5</sup>, Juraj Balkovic<sup>7,8</sup>, Philippe Ciais<sup>9,10</sup>, Marie Dury<sup>11</sup>, Pete Falloon<sup>12</sup>,  
Christian Folberth<sup>7</sup>, Louis François<sup>11</sup>, Tobias Hank<sup>13</sup>, Munir Hoffmann<sup>14,23</sup>, Cesar Izaurrealde<sup>15,16</sup>,  
Ingrid Jacquemin<sup>11</sup>, Curtis Jones<sup>15</sup>, Nikolay Khabarov<sup>7</sup>, Marian Koch<sup>14</sup>, Michelle Li<sup>2,17</sup>, Wenfeng Liu<sup>18,9</sup>,  
Stefan Olin<sup>19</sup>, Meridel Phillips<sup>5,20</sup>, Thomas Pugh<sup>21,22</sup>, Ashwan Reddy<sup>15</sup>, Xuhui Wang<sup>9,10</sup>,  
Karina Williams<sup>12</sup>, Florian Zabel<sup>13</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, University of Chicago, Chicago, IL, USA
3. Department of Computer Science, University of Chicago, Chicago, IL, USA
4. Potsdam Institute for Climate Impact Research, Leibniz Association (Member), Potsdam, Germany
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. Ecosystem Services and Mgm. Prg., International Institute for Applied Systems Analysis, Laxenburg, Austria
8. Department of Soil Science, Comenius University in Bratislava, Bratislava, Slovak Republic
9. Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France
10. Sino-French Institute of Earth System Sciences, Peking University, Beijing, China
11. Unité de Modélisation du Climat et des Cycles Biogéochimiques, University of Liège, Belgium
12. Met Office Hadley Centre, Exeter, United Kingdom
13. Department of Geography, Ludwig-Maximilians-Universität, Munich, Germany
14. Georg-August-University Göttingen, Tropical Plant Production and Ag. Sys. Modelling, Göttingen, Germany
15. Department of Geographical Sciences, University of Maryland, College Park, MD, USA
16. Texas Agrilife Research and Extension, Texas A&M University, Temple, TX, USA
17. Department of Statistics, University of Chicago, Chicago, IL, USA
18. EAWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
19. Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden
20. Earth Institute Center for Climate Systems Research, Columbia University, New York, NY, USA
21. Karlsruhe Institute of Technology, IMK-IFU, 82467 Garmisch-Partenkirchen, Germany
22. School of Geography, Earth and Environmental Science, University of Birmingham, Birmingham, UK
23. Leibniz Centre for Agricultural Landscape Research (ZALF), D-15374 Müncheberg, Germany

## 1 Cultivation Areas

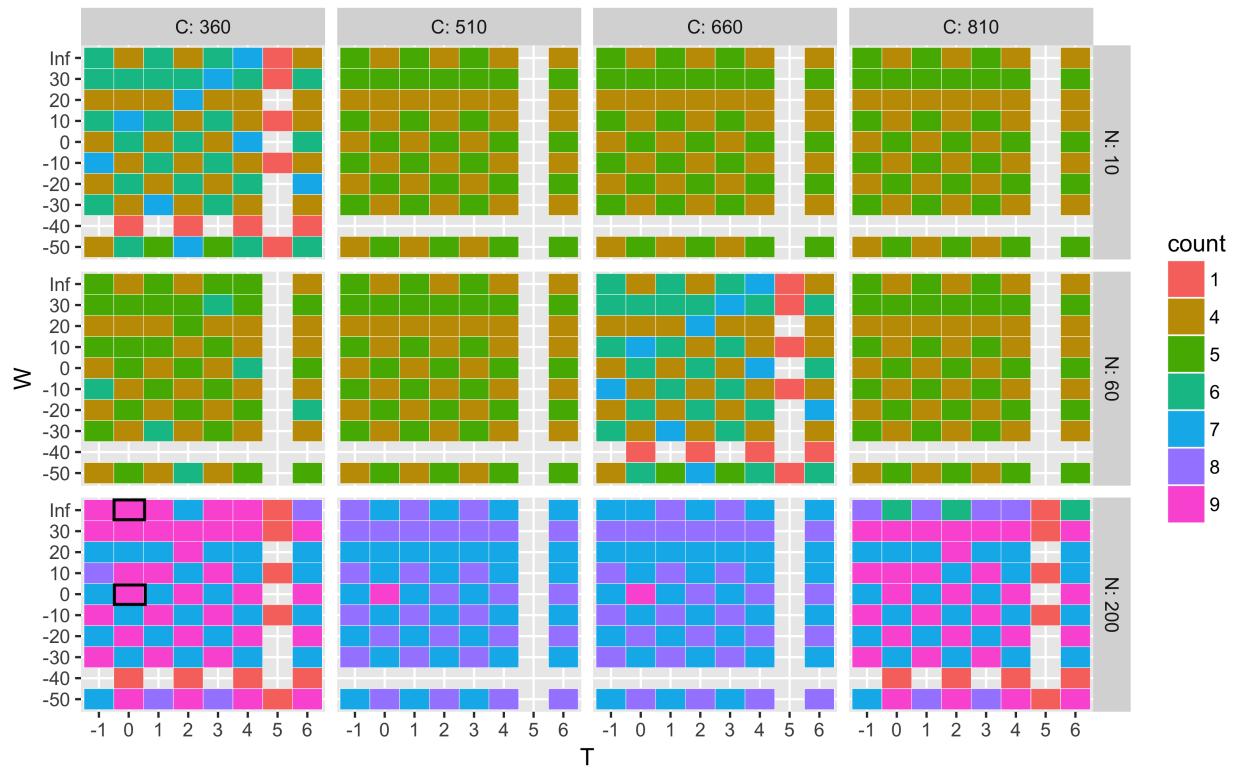


**Figure S1:** Presently cultivated area for irrigated crops in the real world. The blue contour area indicates grid-cells with more than 20,000 hectares of crop cultivated. The gray contour shows area with more than 10 hectares cultivated. Data from the MIRCA2000 data set for maize, rice, and soy. Winter and spring wheat areas are adapted from MIRCA2000 data and sorted by growing season.



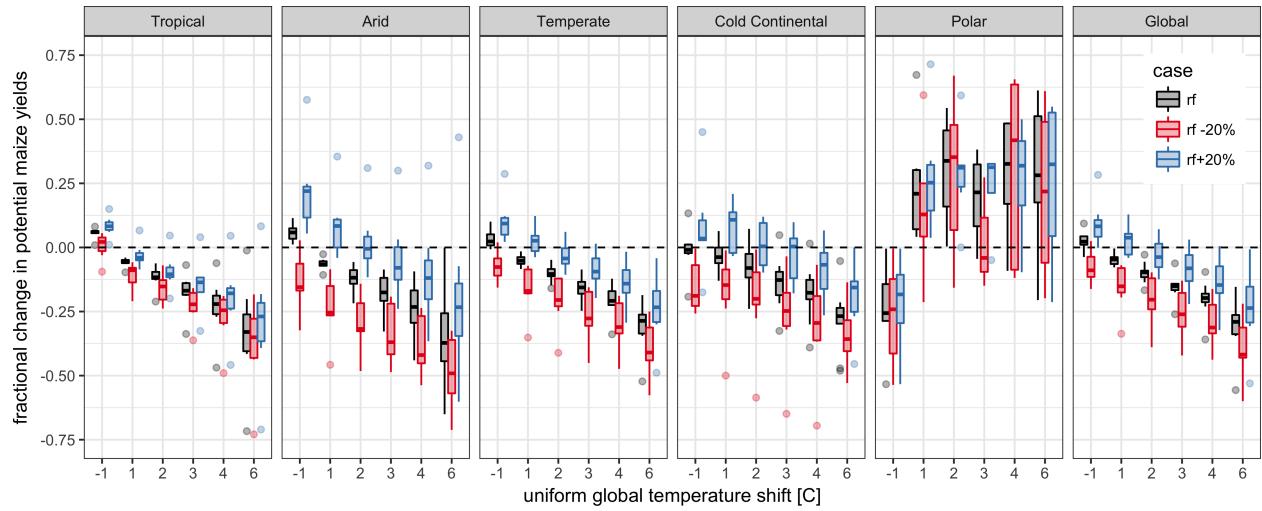
**Figure S2:** Presently cultivated area for rain fed crops in the real world. Conventions as in Figure S1. This figure repeats manuscript Figure 1 for ease of comparison.

## 2 Experiment Simulation Sampling in Variable Space



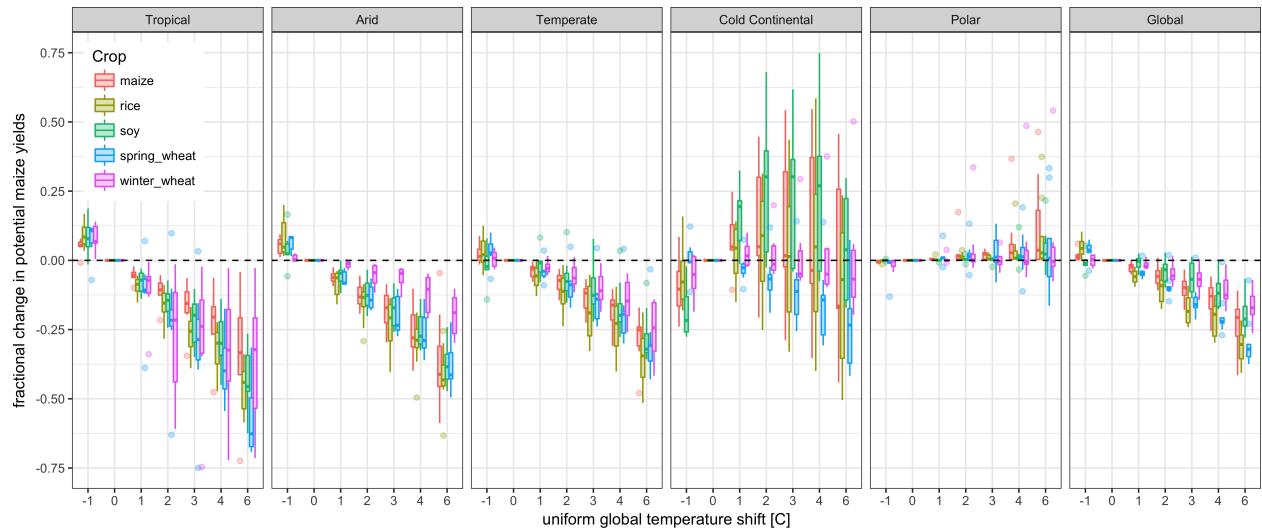
**Figure S3:** Tile heatmap illustrates number of model simulations provided for each of the scenarios in the variable space. The max number is 9, the number of models included in the emulator analysis (excluding three models not included in the emulator analysis). Error calculations are run over scenarios with max number of models (See Figures S25, S26)

### 3 Maize Simulations



**Figure S4:** Maize simulation results. As in Figure 2 in the main text but weighted by actual cultivation area in the real world instead of across all grid cells. Additional figure conventions are the same as Figure 2 in the main text. All other covariates are held constant.

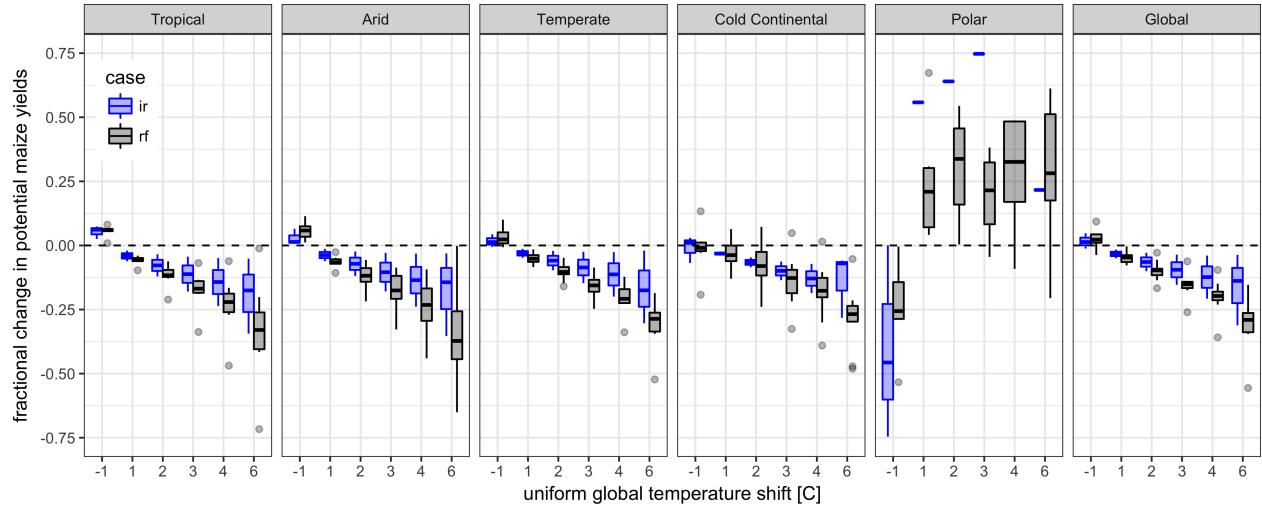
### 4 All Crops Simulations



**Figure S5:** All crop simulation results. As in Figure 2 in the main document except for all models for the rainfed case (without -20 or +20% rainfall as in Figure 2 in main text.) In cold continental regions, soy yields generally increase with warming and spring wheat yields decrease; the other crops are indeterminate across models. All other covariates are held constant.

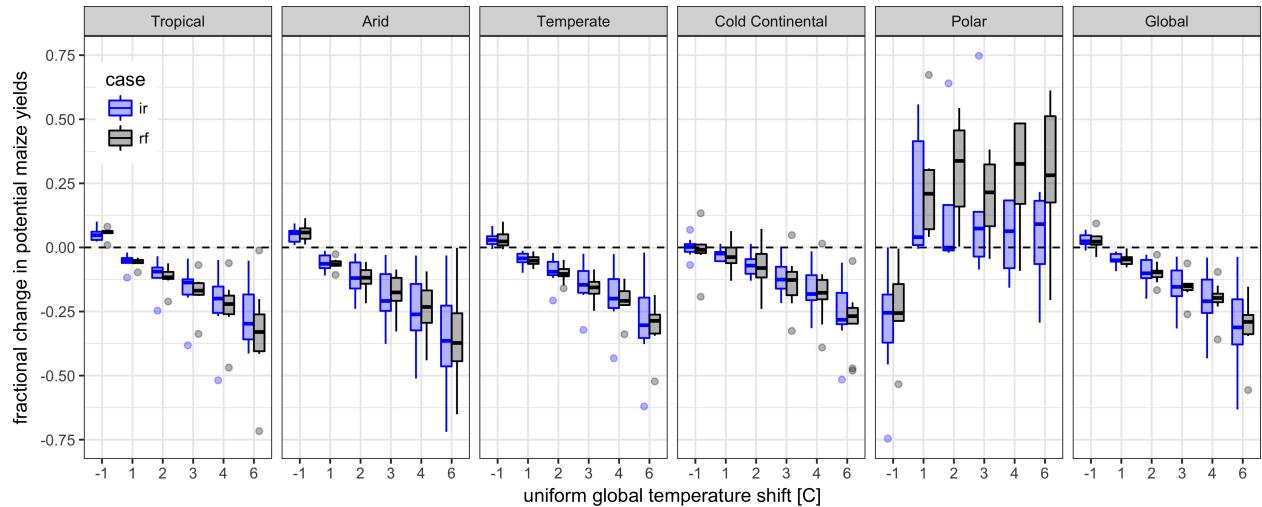
## 5 Maize Simulations

### 5.1 All grid cells



**Figure S6:** Maize simulation results. As in Figure 2 in the main text except comparing rainfed to irrigated maize across all grid cells. All other figure conventions match Figure 2 in the main text.

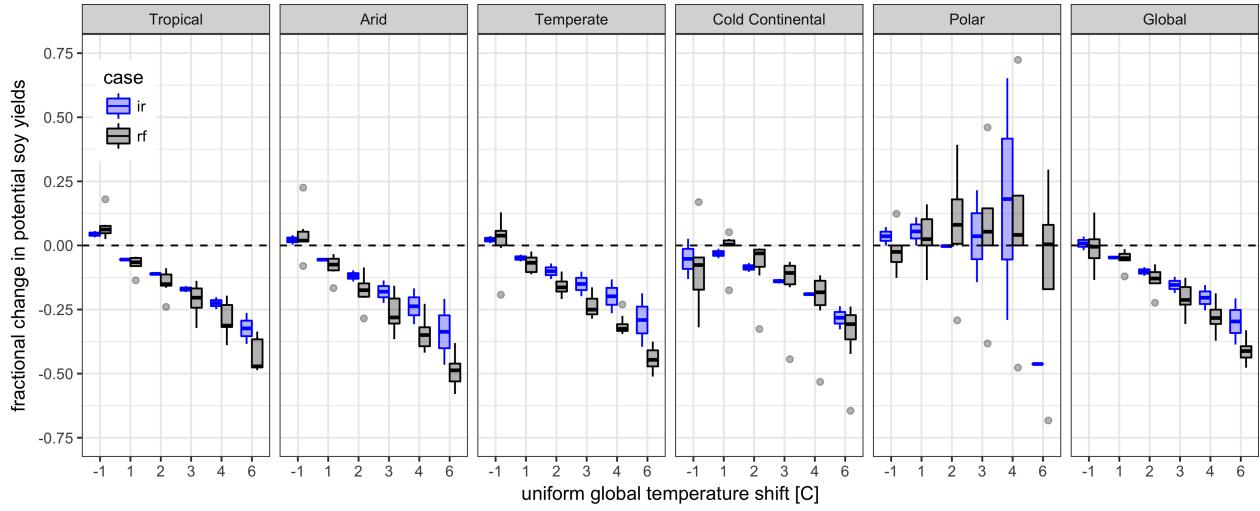
### 5.2 Currently cultivated area



**Figure S7:** Maize simulation results. As in Figure 2 in the main text except comparing rainfed to irrigated maize across currently cultivated hectares. All other figure conventions match Figure 2 in the main text.

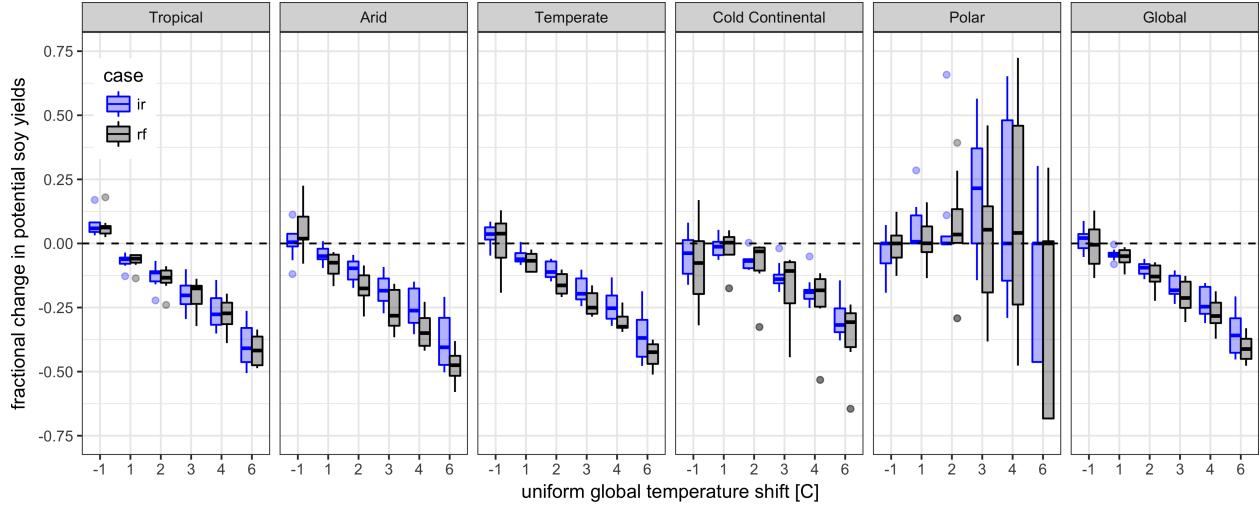
## 6 Soy Simulations

### 6.1 All grid cells



**Figure S8:** Soy simulation results. As in Figure 2 in the main text except comparing rainfed to irrigated soy across all grid cells. All other figure conventions match Figure 2 in the main text.

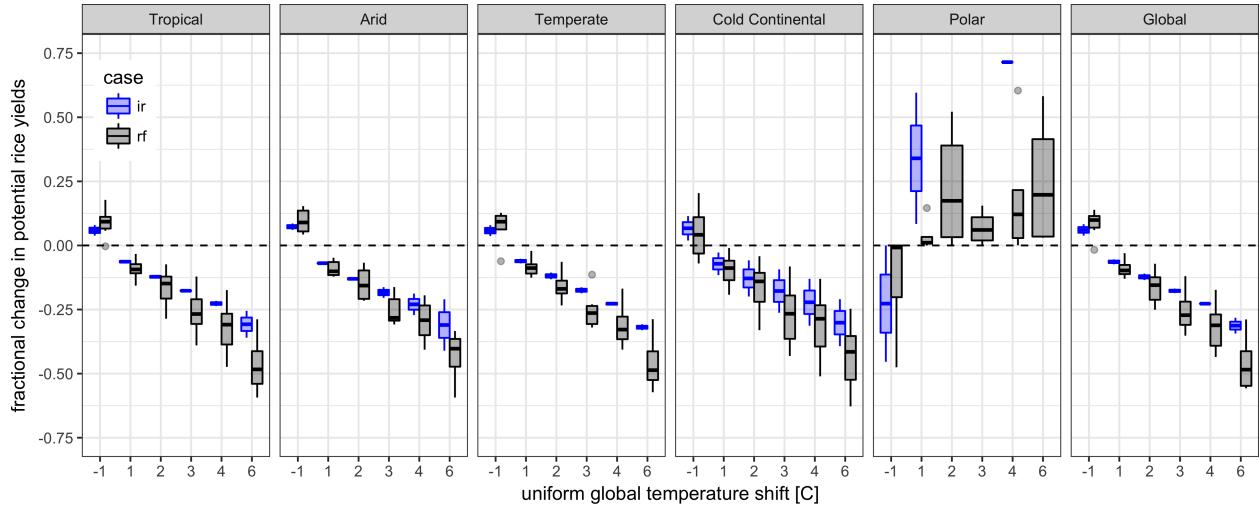
### 6.2 Currently cultivated area



**Figure S9:** Soy simulation results. As in Figure 2 in the main text except comparing rainfed to irrigated soy across currently cultivated hectares. All other figure conventions match Figure 2 in the main text.

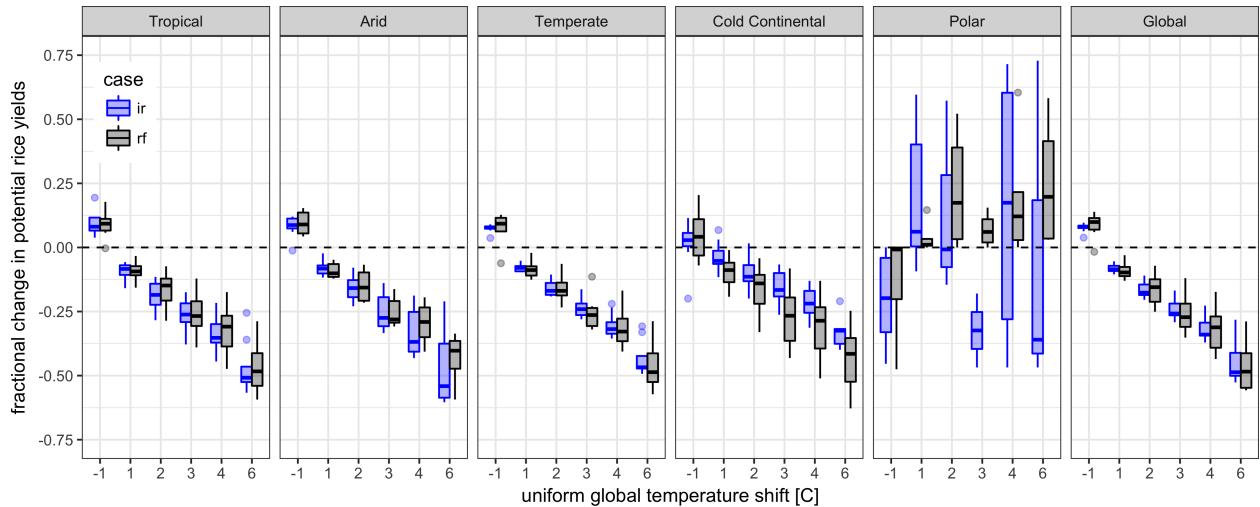
## 7 Rice Simulations

### 7.1 All grid cells



**Figure S10:** Rice simulation results. As in Figure 2 in the main text except comparing rainfed to irrigated rice across all grid cells. All other figure conventions match Figure 2 in the main text.

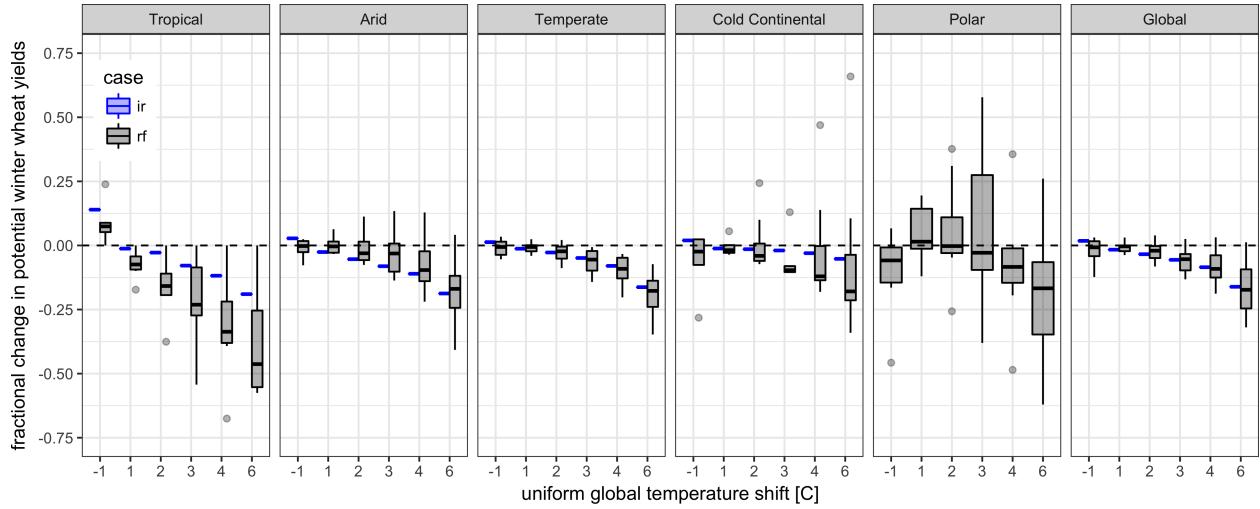
### 7.2 Currently cultivated area



**Figure S11:** Rice simulation results. As in Figure 2 in the main text except comparing rainfed to irrigated rice across currently cultivated hectares. All other figure conventions match Figure 2 in the main text.

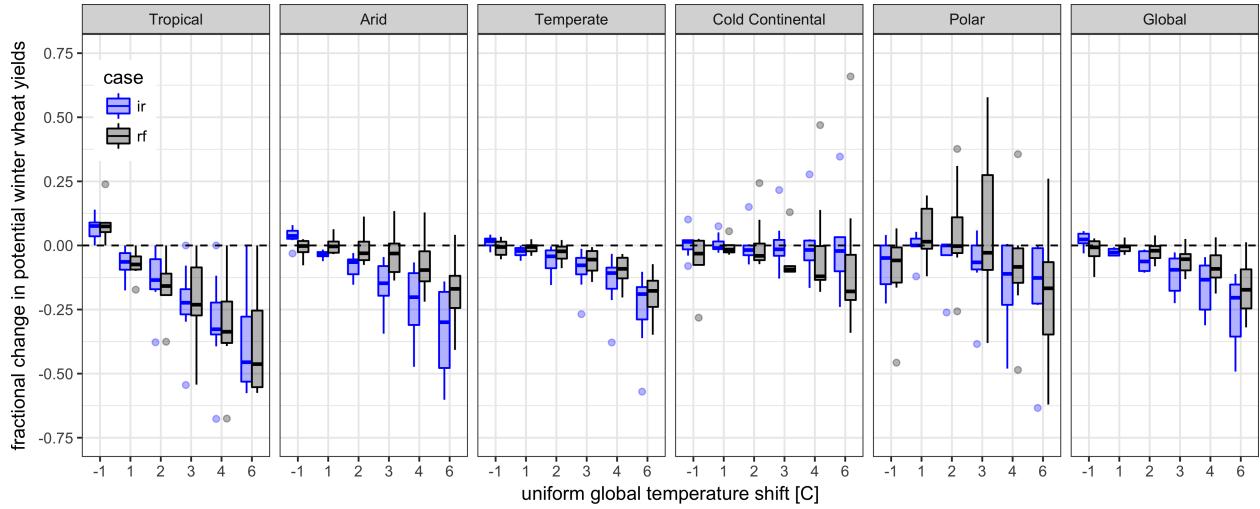
## 8 Winter Wheat Simulations

### 8.1 All grid cells



**Figure S12:** Winter wheat simulation results. As in Figure 2 in the main text except comparing rainfed to irrigated winter wheat across all grid cells. All other figure conventions match Figure 2 in the main text.

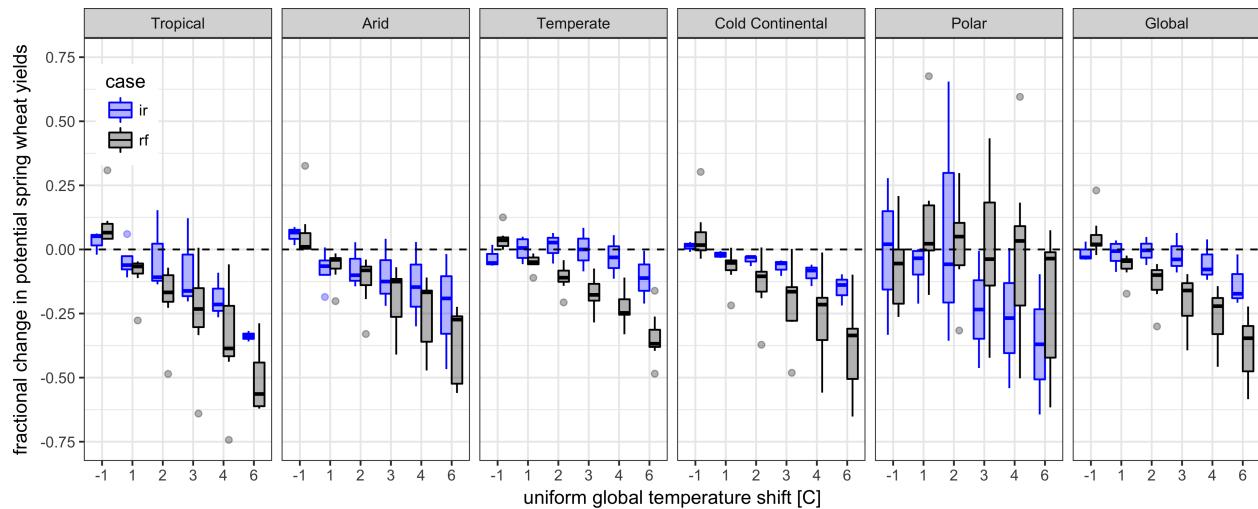
### 8.2 Currently cultivated area



**Figure S13:** Winter wheat simulation results. As in Figure 2 in the main text except comparing rainfed to irrigated winter wheat across currently cultivated hectares. All other figure conventions match Figure 2 in the main text.

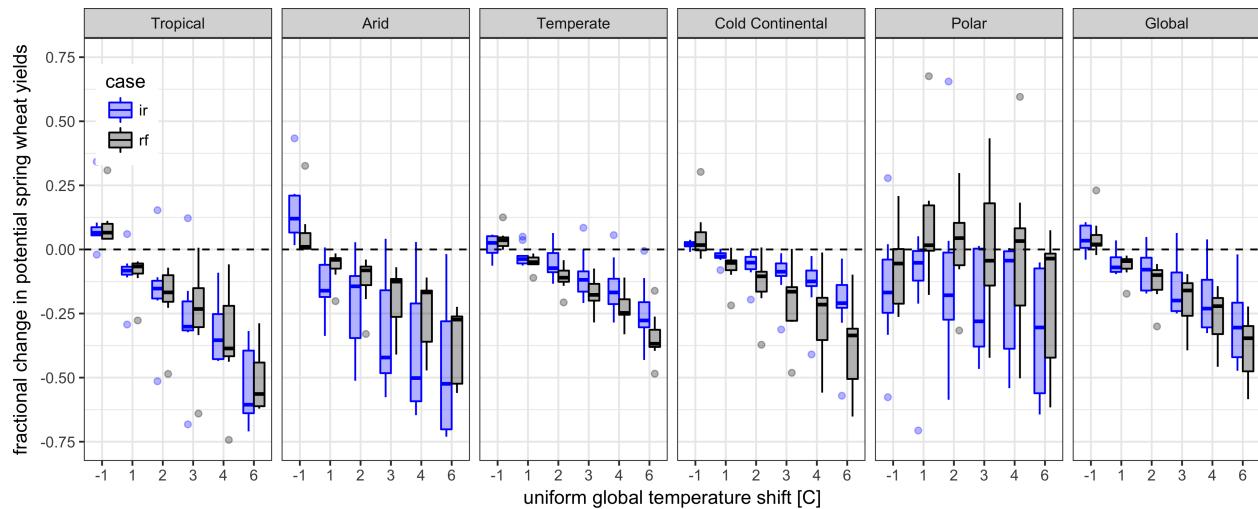
## 9 Spring Wheat Simulations

### 9.1 All grid cells



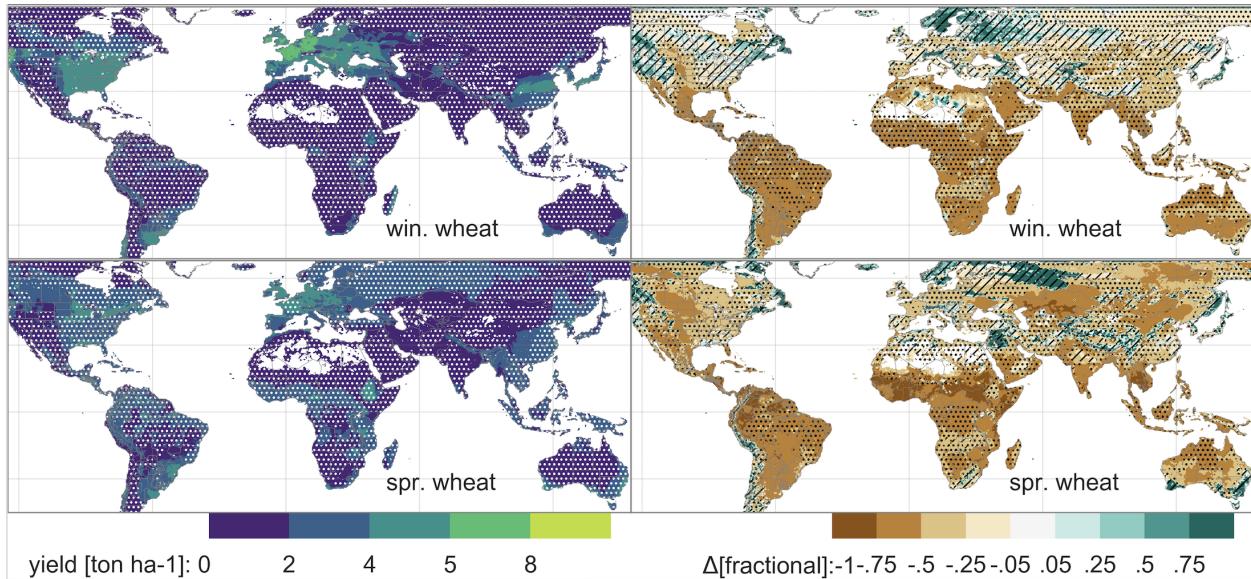
**Figure S14:** Spring wheat simulation results. As in Figure 2 in the main text except comparing rainfed to irrigated spring wheat across all grid cells. All other figure conventions match Figure 2 in the main text.

### 9.2 Currently cultivated area



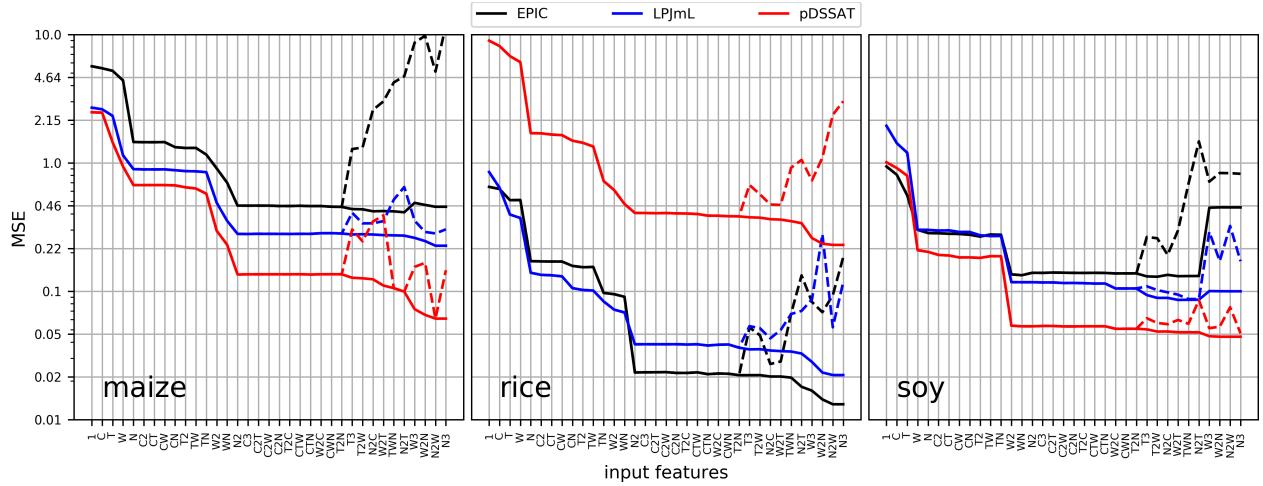
**Figure S15:** Spring wheat simulation results. As in Figure 2 in the main text except comparing rainfed to irrigated spring wheat across currently cultivated hectares. All other figure conventions match Figure 2 in the main text.

## 10 Wheat Simulations



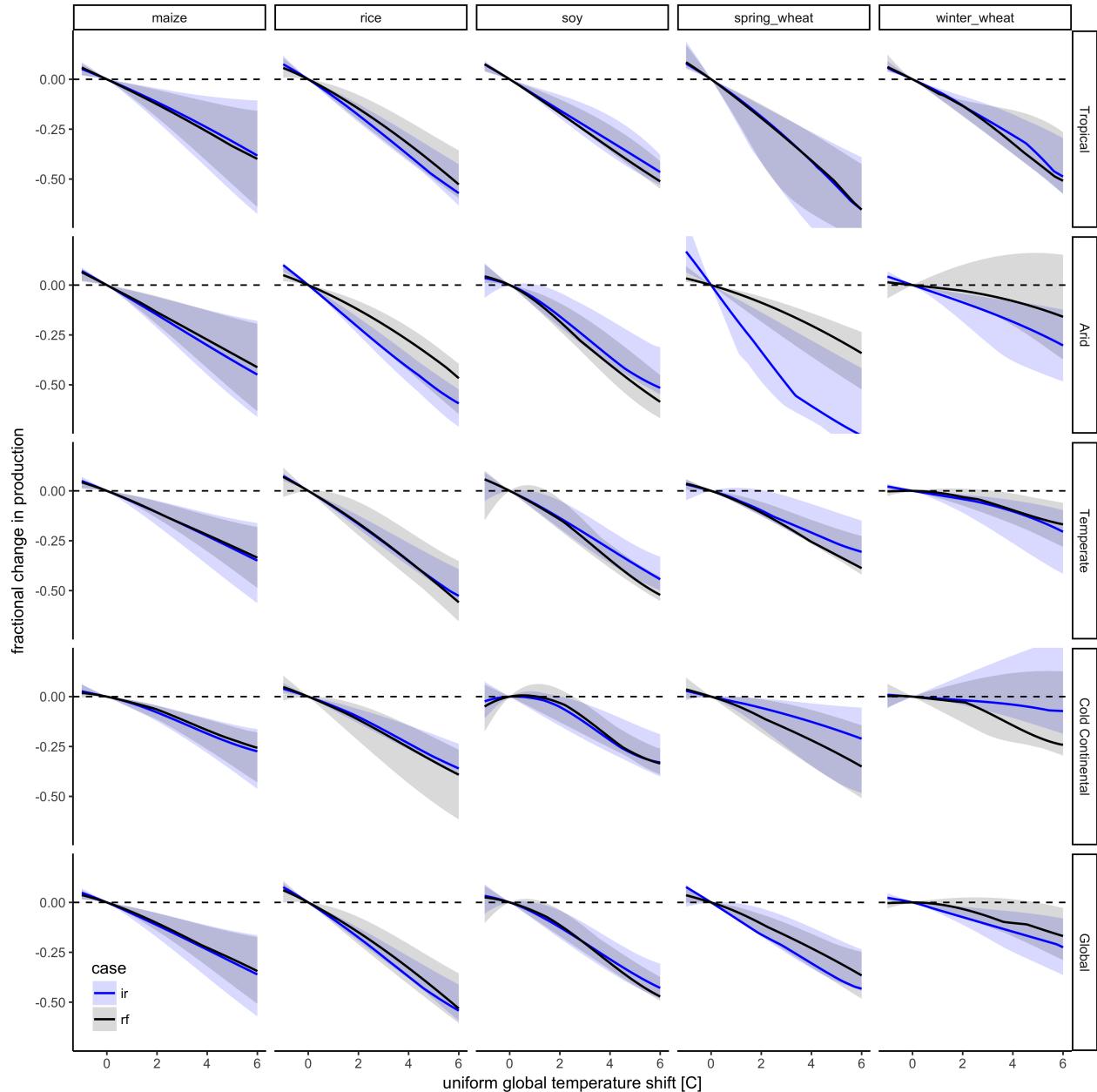
**Figure S16:** Illustration of the spatial pattern of potential yields and potential yield changes in the GGCMI Phase II ensemble, for wheat. All figure conventions follow Figure 3 in the main text. Wheat model results in cold areas, where yield impacts are on average positive, also have the highest uncertainty. Wheat is also somewhat exceptional in that also less impact in temperature and arid regions. The more complicated phenological development of winter wheat when compared to other crops is a potential source of the higher level of model disagreement.

## 11 Emulator



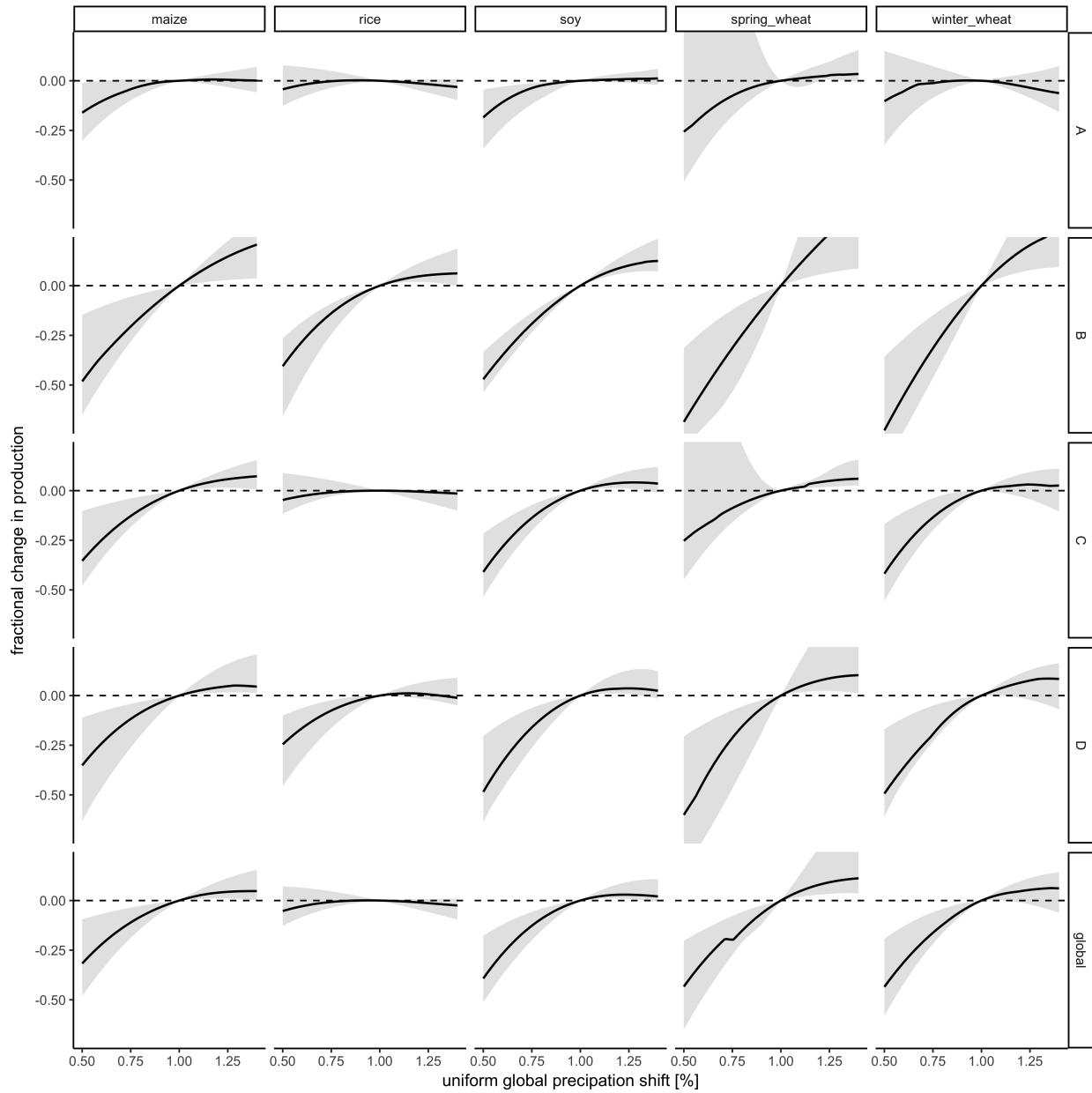
**Figure S17:** Mean Squared Error (MSE) from the cross-validation process during feature selection. The emulate yield for maize rice and soy are compared to the simulated values for the same model in all grid cells where crops are actually cultivated, weighted by area in each grid cell. The X- axis indicates terms included in the model at each step progressively where T = temperature, T2 = temperature<sup>2</sup>, TW = temperature \* water and so on. The terms that did not reduce the aggregate error (horizontal lines) are not included in the final model. Some terms that did not reduce the aggregate error are included if a higher order version of that term provided a decrease in mean squared error (i.e. temperature cubed cannot be included without the temperature squared term, and the linear temperature term). Even though the models exhibit different absolute levels of error, all three models agree remarkably well on feature importance indicated by the locations where the lines degenerate and where they stay horizontal. Solid lines indicate L1 normalization and dashed lines indicate L2 normalization of yield outputs. Colors indicated different models (three fully-sampled simulation sets.)

## 11.1 Emulated damage function for Temperature



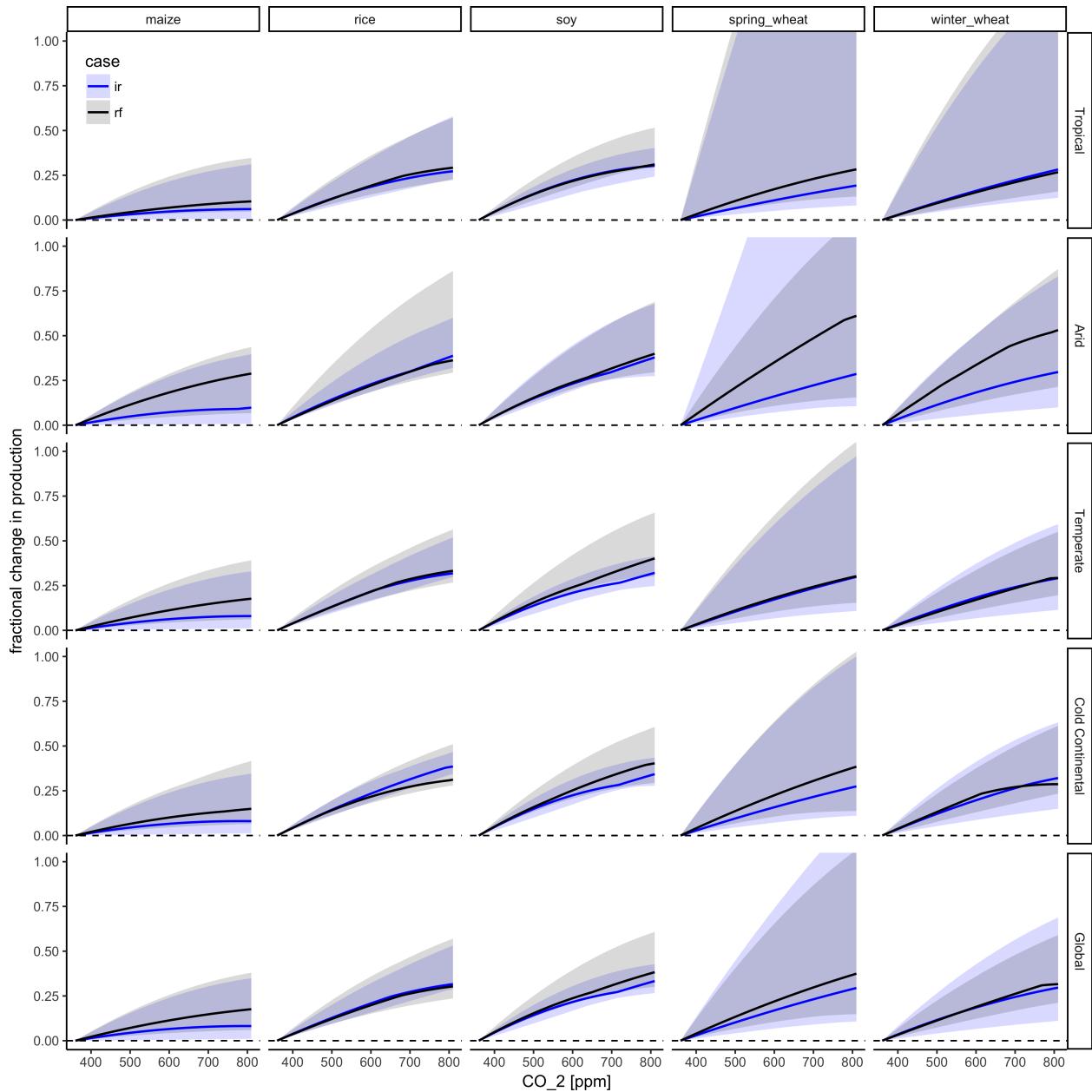
**Figure S18:** Multi-model ensemble spread in sensitivity to changes in the temperature dimension. 5%, median, and 95% percentile emulated damage function for currently cultivated areas. Irrigated and rain-fed crops shown. PROMET and JULES removed from ensemble. All other covariates held constant. Uniform temperature shift should not be interpreted as a realistic climate change.

## 11.2 Emulated damage function for Water



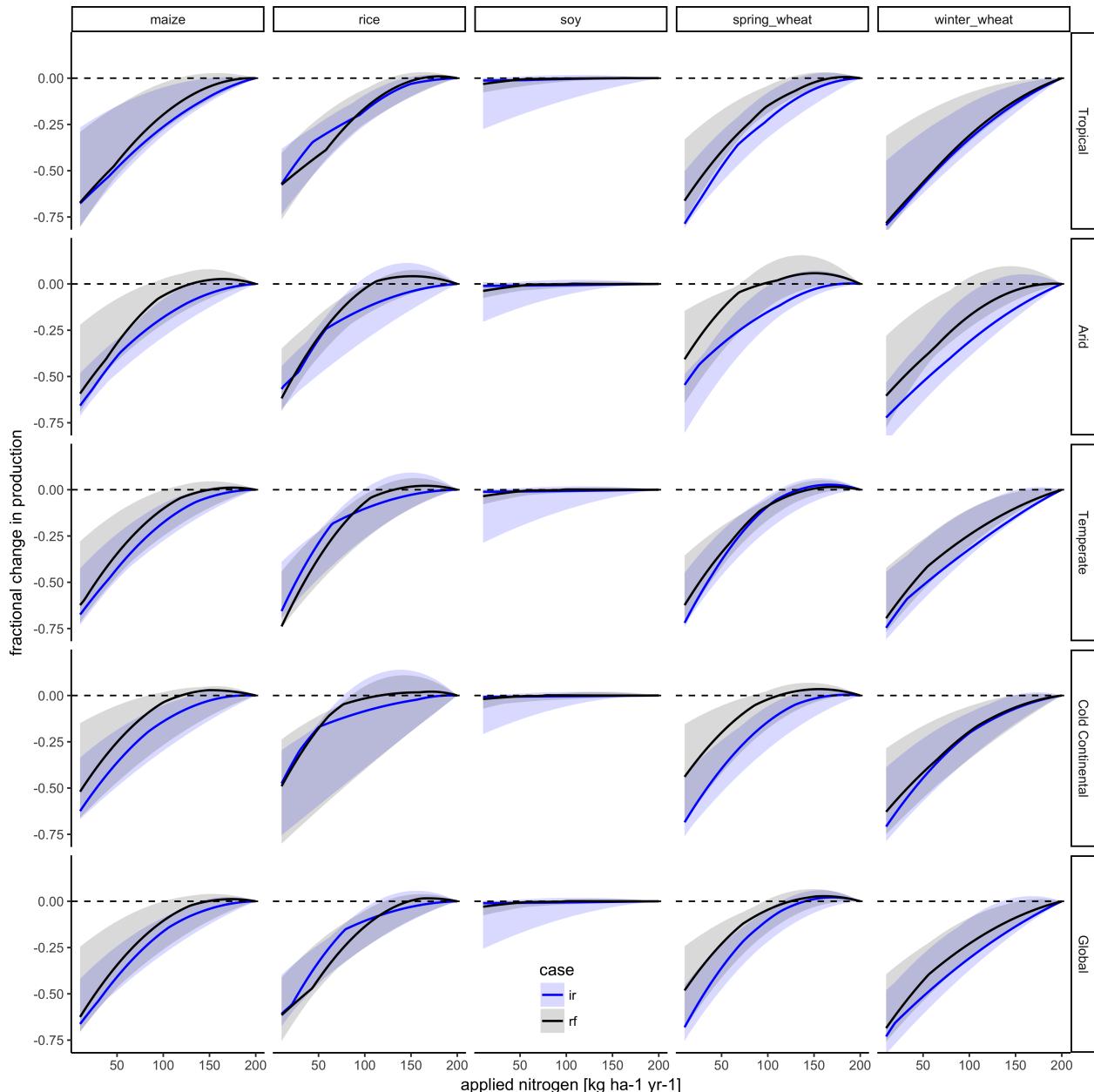
**Figure S19:** Multi-model ensemble spread in sensitivity to changes in the water dimension. 5%, median, and 95% percentile emulated damage function for currently cultivated areas. PROMET and JULES removed from ensemble. All other covariates held constant. Uniform precipitation shift should not be interpreted as a realistic climate change.

### 11.3 Emulated damage function for Carbon



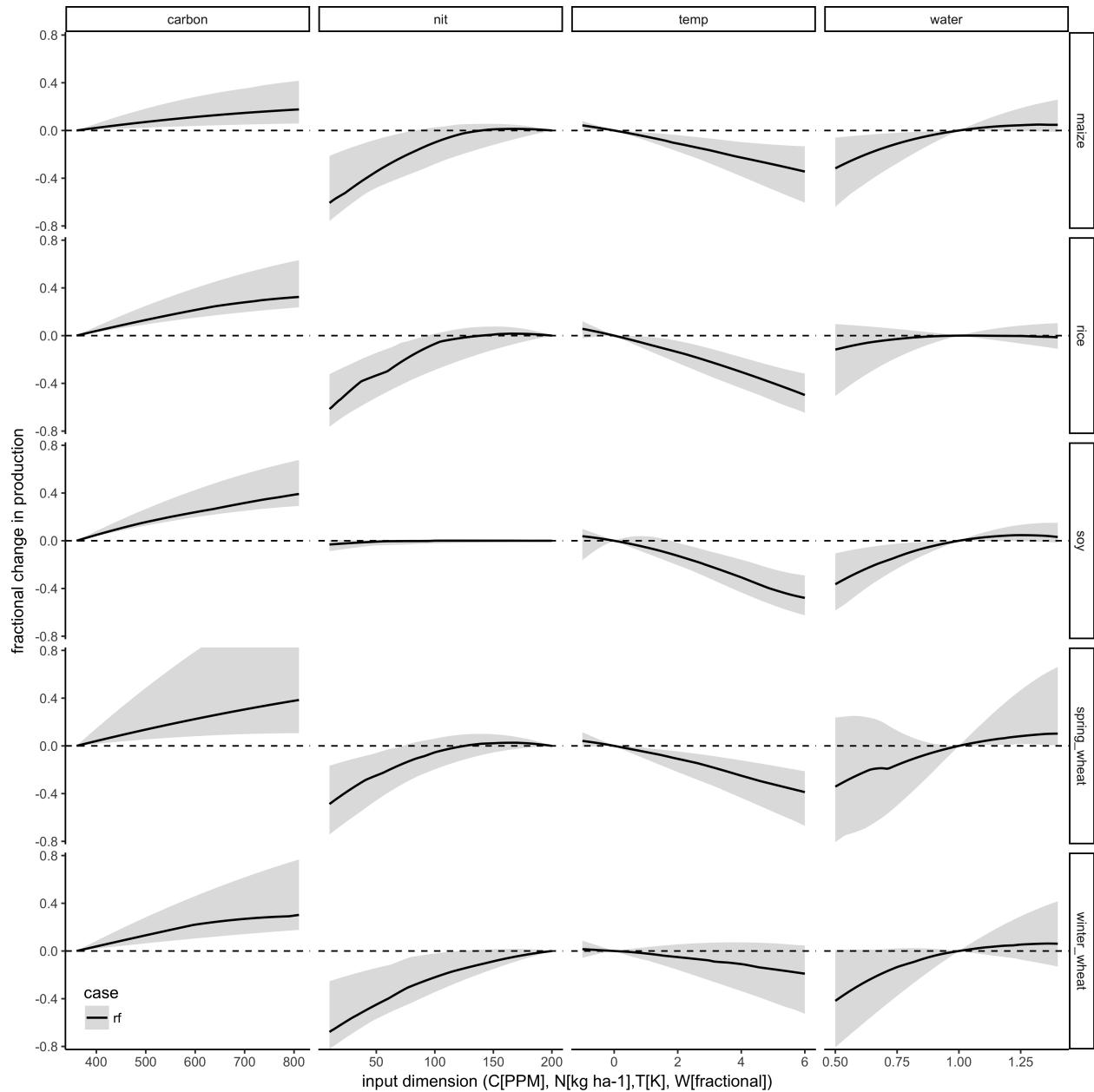
**Figure S20:** Multi-model ensemble spread in sensitivity to changes in the carbon dimension. 5%, median, and 95% percentile emulated damage function for currently cultivated areas. Irrigated and rain-fed crops shown. PROMET and JULES removed from ensemble. All other covariates held constant.

## 11.4 Emulated damage function for Nitrogen

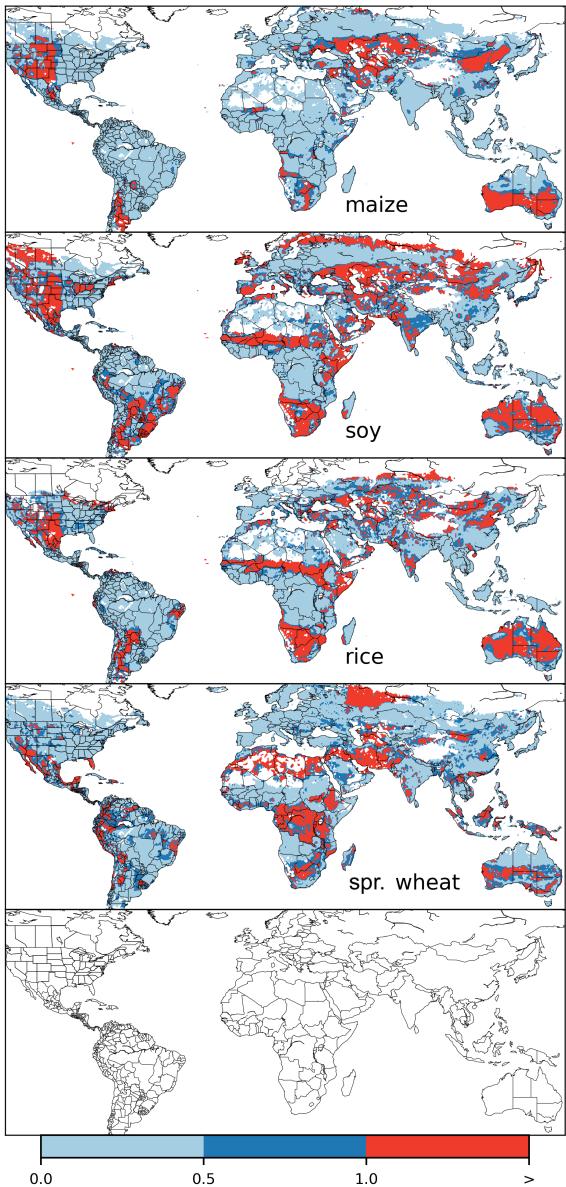


**Figure S21:** Multi-model ensemble spread in sensitivity to changes in the nitrogen dimension. 5%, median, and 95% emulated damage function for currently cultivated areas. Irrigated and rain-fed crops shown. PROMET and JULES removed from ensemble. All other covariates held constant in each case.

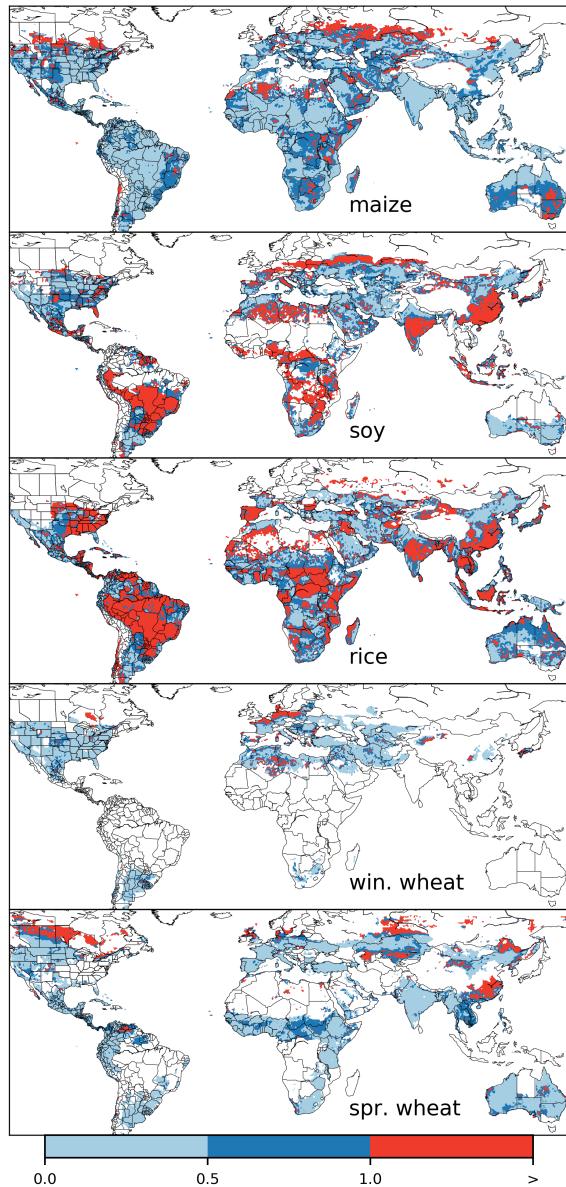
## 11.5 Emulated damage function for all dimensions



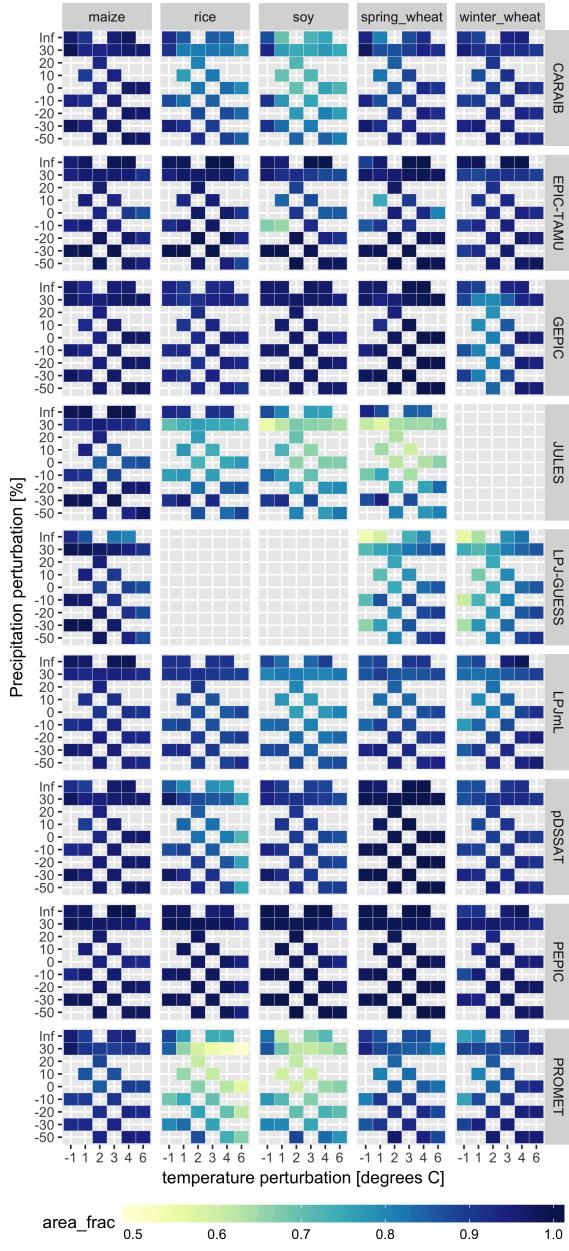
**Figure S22:** Multi-model ensemble spread in sensitivity to changes in all four dimensions for rain-fed crops at the global level. 5% median and 95% percentile emulated damage function for currently cultivated areas. PROMET and JULES removed from ensemble. All other covariates held constant in each case.



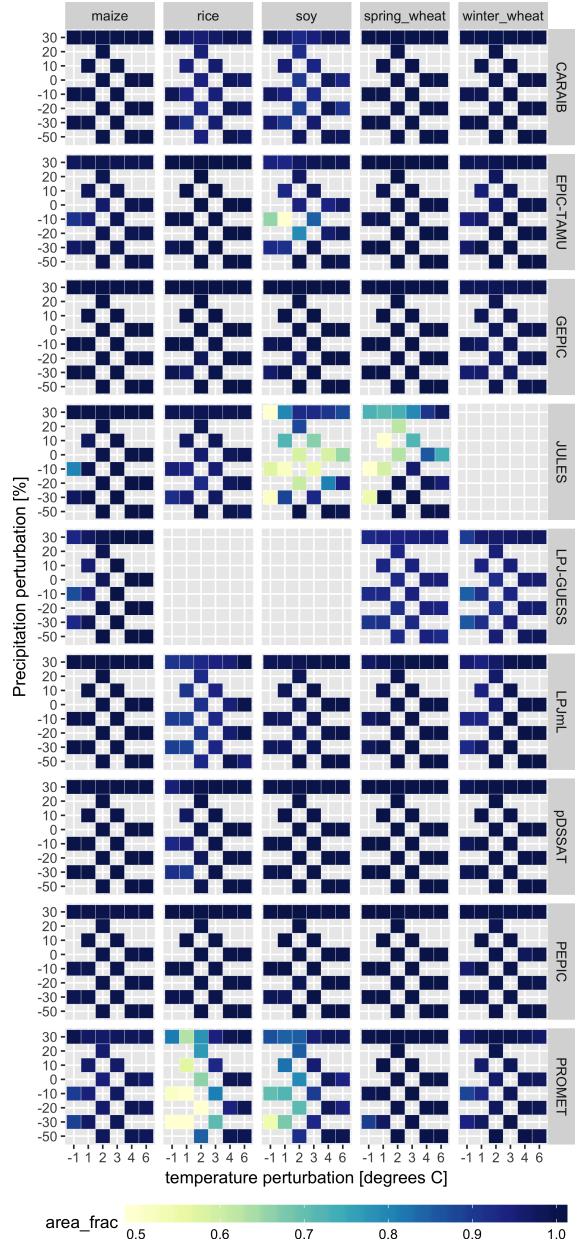
**Figure S23:** Illustration of our test of emulator performance, applied to the JULES model for the T +4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error  $e$ , where  $e$  less than 1 means that emulator error exceeds the multi-model standard deviation. White areas are those where crops are not simulated by this model. Models differ in their areas omitted, meaning the number of samples used to calculate the multi-model standard deviation is not spatially consistent in all locations.



**Figure S24:** Illustration of our test of emulator performance, applied to the PROMET model for the T +4 scenario for rain-fed crops. Contour colors indicate the normalized emulator error  $e$ , where  $e$  less than 1 means that emulator error exceeds the multi-model standard deviation. White areas are those where crops are not simulated by this model. Models differ in their areas omitted, meaning the number of samples used to calculate the multi-model standard deviation is not spatially consistent in all locations.



**Figure S25:** The fraction of grid cells (across all grid cells) with normalized emulation error less than 1 for the  $\text{CO}_2=360 \text{ ppm}$  and  $200 \text{ kg N ha}^{-1} \text{ yr}^{-1}$  case for the temperature and precipitation perturbations scenarios provided by all 9 models included in the emulator analysis. This is in contrast to the fraction of currently cultivated hectares shown in the C360 case in the manuscript and for the C810 case show in the supplemental material. The emulator is marginally more successful over currently cultivated areas than over all grid cells in general.



**Figure S26:** The fraction of currently cultivated hectares with normalized emulation error less than 1 (blue colors contours in Figure A2) for the  $\text{CO}_2=810 \text{ ppm}$  and  $200 \text{ kg N ha}^{-1} \text{ yr}^{-1}$  case for the temperature and precipitation perturbations scenarios provided by all 9 models included in the emulator analysis. See Equations A1 and A2 for normalized error calculation. The yield response is generally easy to emulate over currently cultivated areas (dark blue and light blue).