IOP Publishing Environ. Res. Lett. 12 (2017) 075007 https://doi.org/10.1088/1748-9326/aa788c

Environmental Research Letters

OPEN ACCESS

RECEIVED
8 November 2016

REVISED
8 June 2017

ACCEPTED FOR PUBLICATION
9 June 2017

PUBLISHED
10 July 2017

LETTER

Simulated vs. empirical weather responsiveness of crop yields: US evidence and implications for the agricultural impacts of climate change

Malcolm N Mistry[1,2,4], Ian Sue Wing[3] and Enrica De Cian[2]

1 Department of Economics, Università Ca' Foscari, Venice, Italy
2 Fondazione Eni Enrico Mattei (FEEM), Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Venice, Italy
3 Department of Earth and Environment, Boston University, Boston, MA, United States of America
4 Author to whom any correspondence should be addressed.

E-mail: malcolm.mistry@feem.it, malcolm.mistry@unive.it

Keywords: climate change impacts, crop yields, global gridded crop models, ISI-MIP

Supplementary material for this article is available online

Abstract

Global gridded crop models (GGCMs) are the workhorse of assessments of the agricultural impacts of climate change. Yet the changes in crop yields projected by different models in response to the same meteorological forcing can differ substantially. Through an inter-method comparison, we provide a first glimpse into the origins and implications of this divergence—both among GGCMs and between GGCMs and historical observations. We examine yields of rainfed maize, wheat, and soybeans simulated by six GGCMs as part of the Inter-Sectoral Impact Model Intercomparison Project-Fast Track (ISIMIP-FT) exercise, comparing 1981–2004 hindcast yields over the coterminous United States (US) against US Department of Agriculture (USDA) time series for about 1000 counties. Leveraging the empirical climate change impacts literature, we estimate reduced-form econometric models of crop yield responses to temperature and precipitation exposures for both GGCMs and observations. We find that up to 60% of the variance in both simulated and observed yields is attributable to weather variation. A majority of the GGCMs have difficulty reproducing the observed distribution of percentage yield anomalies, and exhibit aggregate responses that show yields to be more weather-sensitive than in the observational record over the predominant range of temperature and precipitation conditions. This disparity is largely attributable to heterogeneity in GGCMs' responses, as opposed to uncertainty in historical weather forcings, and is responsible for widely divergent impacts of climate on future crop yields.

1. Introduction

Agriculture, particularly cultivation of field crops, is weather dependent and exposed to meteorological shifts (Gornall *et al* 2010, Moore and Lobell 2015), making it especially vulnerable to adverse effects of climate change (IPCC 2014). The specter of declining yields of maize, wheat, soybeans and other food staples with exposure to high temperature and low precipitation extremes arises from two lines of evidence (Moore and Lobell 2015, Lobell *et al* 2011, Porter *et al* 2014, Müller *et al* 2015, Lobell and Asseng 2017). First, the empirical climate change economics literature estimates reduced-form responses of yields to weather shocks using historically observed production, harvested area, temperature and precipitation in many locations across multiple years (e.g. Lobell *et al* 2011, Porter *et al* 2014, Schlenker and Lobell 2010, Tack *et al* 2015). Second, process-based crop models simulate the detailed influences on plant growth of a wide array of weather variables, plant genotypes, environmental factors such as the carbon dioxide (CO_2) fertilization effect (CFE), soil quality or pests, and agronomic adaptations such as irrigation, fertilizer application, and the timing of planting and harvesting (Elliott *et al* 2015, Bassu *et al* 2014, Rosenzweig *et al* 2014).

Whereas the geographic domain of empirical studies is often limited to individual countries or regions with a sufficient number of historical observations[5], global gridded crop models (GGCMs) simulate the growth of field crops worldwide under different climatic conditions projected by earth system models (ESMs) (see Deryng *et al* 2011, Rosenzweig *et al* 2014 and Elliott *et al* 2015 for further discussion), resulting in a comprehensive picture of the effects of climate change on crop yields.

Confidence in GGCMs' simulated agricultural impacts turns on the ability of models to accurately capture the myriad interacting meteorologically-driven processes that determine yields (Bassu *et al* 2014). GGCMs' representations of plant growth dynamics rely on numerous parameters that must be calibrated, but whose values are uncertain and may vary geographically in ways that are poorly constrained (Rosenzweig *et al* 2014, Jones *et al* 2016). Validation typically involves statistical evaluation of GGCMs' ability to reproduce point estimates of yields at different locations, for example at field trial sites or over spatially aggregated production regions under year-to-year variation in weather conditions (for excellent recent examples, see Morell *et al* 2016, Müller *et al* 2017). However, comparatively little attention has been paid to how the *response* of GGCMs-simulated yields to meteorological forcings compare with the weather sensitivity of yields observed in observed agricultural systems. Early studies focused on a single crop model (Lobell and Burke 2010, Watson *et al* 2015), and recent availability of extensive multi-model cross-section/time-series crop yields datasets generated by GGCM intercomparison exercises have facilitated reduced-form statistical emulation of single (Oyebamiji *et al* 2015) or multiple-GGCM (Blanc and Sultan 2015, Blanc 2017) simulations, for one or more crops (Blanc 2017). However, except for Lobell and Asseng 2017 and Schauberger *et al* 2017, such emulators do not appear to have been used for *diagnostic* purposes. It is this gap that we address here[6], by comparing the responses of process simulations with those of econometric models trained on observations. Our strategy is to elucidate and compare the aggregate responses of observed and GGCM-simulated yields to observed and ESM-simulated temperature and precipitation under current climatic conditions. We pose six key questions:

Q.I How well do the outputs of GGCM hindcast simulations match historically observed yields?

Q.II Do GGCMs reproduce the correlations between yields and adverse (i.e. high temperature and low precipitation) weather extremes seen in the observational record?

Q.III How similar are GGCM-simulated and observed yield responses, under not only adverse extremes, but the full range of weather conditions over crops' growing seasons?

Q.IV Do differences between GGCMs and observations in the weather-responsive component of yields originate in divergent meteorological forcings (i.e. differences in temperature and precipitation exposures between weather observations and ESM historical simulations), versus divergence in GGCMs' simulated responses and observed crop responses to these forcings?

Q.V To which characteristics of GGCMs can the divergence between simulated and observed responses be attributed?

Q.VI What do simulated and observed response functions imply for the impacts of climate change-driven shifts in temperature and precipitation on future United States (US) crop yields?

To provide answers we statistically extract and compare the responses of yield to weather shocks for two sets of data that span the same temporal and spatial domain: rainfed maize, wheat and soybeans in the coterminous US over the period 1981–2004. For crop models we use the outputs of runs of six GGCMs fielded by the Inter-Sectoral Impact Model Intercomparison Project Fast-Track (ISIMIP-FT) exercise (Warszawski *et al* 2013, Rosenzweig *et al* 2014, Frieler *et al* 2015), together with their bias-corrected ESM-simulated meteorological forcings (Hempel *et al* 2013). For historical observations, we use US Dept. of Agriculture (USDA) multi-decadal time series of production and harvested area for about 1000 predominantly rainfed counties (whose areal extents are comparable to GGCMs' grid cells across US farm states), matched to high-frequency temperature and precipitation exposures from a historical weather dataset.

The rest of the paper is organized as follows. Section 2 discusses our data and elaborates the methods we use to answer questions I–VI. A discussion of the results is provided in section 3. We summarize our findings with the associated caveats and recommendations for future research in section 4.

2. Methods

Our data consist of m unbalanced panel datasets of maize, wheat and soybean yields (Y) that are either observed or modeled at i areal units over t years, matched with observed or simulated daily temperature (T) and precipitation (P) over the growing season for

[5] For examples, see Iglesias *et al* (2000) for Spain, Lobell and Burke (2010) for US counties, Lobell *et al* (2012) for India, Schlenker and Lobell (2010) for Sub-Saharan Africa.

[6] Whereas Lobell and Asseng (2017) focus on identifying systematic differences between process-based and statistical methods, Schauberger *et al* (2017) address the yield losses in maize, soybeans and winter wheat (rainfed and irrigated) attributable to high-temperature induced mechanisms.

the same locations and periods. Historical crop yields were computed from 1981–2004 county production and harvested area records from the USDA National Agricultural Statistics Service's Quickstats 2.0 database, which provides survey data[7]. Historical weather exposures are calculated from the Parameter-elevation Regressions on Independent Slopes Model (PRISM)[8] forcing files, which are daily meteorological fields on a 2.5 arcmin (~4 km) grid that we spatially interpolate to county boundaries. Simulated 1981–2004 yields on a 0.5° grid were taken from the ISMIP-FT ESGF node[9] for six GGCMs: GEPIC (Liu *et al* 2007), GAEZ-IMAGE (Bouwman *et al* 2006), LPJ-GUESS (Sitch *et al* 2003), LPJmL (Bondeau *et al* 2007, Sitch *et al* 2003), pDSSAT (Elliott *et al* 2013, Jones *et al* 2003) and PEGASUS (Deryng *et al* 2011). Model runs are forced by historical bias-corrected meteorology simulated by the HadGEM2-ES climate model (Jones *et al* 2011) at the same resolution. Further details of the data and models are given in the supplementary information (SI) available at stacks.iop.org/ERL/12/075007/mmedia.

Several factors complicate assessment of GGCMs' skill in reproducing the spatial and temporal patterns of observed yields (Q.I). GAEZ-IMAGE and LPJ-GUESS simulate potential yields while the remaining models simulate actual yields[10], and models are calibrated using historical yields from different sources, whereas others are not calibrated (see Rosenzweig *et al* 2014 SI for further details). For consistency, we characterize the distribution of the differences between the cross-section/time-series yield anomalies of GGCMs ($m = g$) and observations ($m =$ USDA), ${}^*Y_{i,t,\mathrm{GGCMs}} - {}^*Y_{i,t,\mathrm{USDA}}$. Anomalies are defined as fractional deviations from the de-trended long-run mean yield in each location, ${}^*Y_{i,t,g} = Y_{i,t,g}/\overline{Y}_{i,g} - 1$. If ${}^*Y_{i,t,g}$ and ${}^*Y_{i,t,\mathrm{USDA}}$ are similar, then we would expect the probability density function (PDF) of the anomaly difference to be sharply peaked at zero mean.

Our computed anomalies facilitate comparison of the covariation between yields and adverse weather (Q.II). Using a fixed annual growing season[11], we calculate the days of each GGCM (USDA) grid cell's (county's) exposure to j intervals of temperature, ξ_j^T, and k intervals of precipitation, ξ_k^P (see supplementary section S4). We then group grid cells by county, and for both simulations and observational datasets compute the county-level temporal correlations between de-trended yield, ${}^*Y_i$, and the extreme temperature and precipitation bins ($j: T > 30\,°\mathrm{C}$, $k: P \leq 5$ mm).

Taking this analysis one step further, we quantify the potentially nonlinear influence of climate on yields (Q.III) using a semi-parametric cross-section/time-series regression model, following the empirical climate-change impacts literature (Schlenker and Roberts 2006, 2009, Deschênes and Greenstone 2007, 2012, Lobell *et al* 2011, Ortiz-Bobea 2013, Wing *et al* 2015, Burke and Emerick 2016, Schauberger *et al* 2017). For each dataset we specify the dependent variable as the natural logarithm of annual yield (y), and the explanatory variables as a vector of location-specific effects (μ, which capture the influence of unobserved time-invariant local characteristics such as topography and soils), a time-varying function, $f(t)$, which captures the influence of unobserved time-varying shocks, and the vectors of weather exposure covariates ξ_j^T and ξ_k^P described above, and append a random disturbance term, ε:

$$y_{i,t,m} = \mu_i + f(t) + \Sigma_j \beta_{j,m}^T \xi_{j,i,t}^T + \Sigma_k \beta_{k,m}^P \xi_{k,i,t}^P + \varepsilon_{i,t,m} \quad (1)$$

We estimate equation (1) via ordinary least squares on the observational dataset of USDA yield and PRISM weather, the six datasets of GGCM yield outputs and ESM weather inputs, and multi-model panel consisting of the combined inputs and outputs of the six GGCMs[12]. Specifying the function $f(\cdot)$ involves tradeoffs in temporal and spatial flexibility: time effects ($f(t) = \tau_t$) capture the secular influence of year-to-year shocks common to all counties, while geographic variation in trending influences (e.g. input prices, technology adoption, management practices) can be captured by state-specific linear time trends ($f(t) = \lambda_s t$)[13].

Of interest in equation (1) are the estimated parameters $\boldsymbol{\beta}_m^T$ and $\boldsymbol{\beta}_m^P$, vectors of semi-elasticities that capture the average percentage shift in county-level ($m =$ USDA) and grid-level ($g \in m$) yields relative to their conditional mean quantities in response to an additional day in a given interval of temperature or precipitation. Each element of these vectors captures the marginal effect of an additional day of exposure within the corresponding interval (e.g. the average effect of one more day with 25 °C–27 °C versus >30 °C average temperature). Together, the elements flexibly trace out the aggregate response of yields to temperature and precipitation as piecewise linear splines. The latter are statistically identified from the contemporaneous covariation

[7] http://quickstats.nass.usda.gov/ (accessed on 13 February 2017).
[8] PRISM daily data (1981~2004) accessed from www.ocs.orst.edu/prism/ on 13 February 2017.
[9] https://esg.pik-potsdam.de/search/isimip-ft/.
[10] Rosenzweig *et al* (2014) define potential yields as 'unlimited by nutrient or management constraints and without calibration of growth parameter to reproduce historical yields'.
[11] For both simulated and historical datasets, we define the growing season as April–August (AMJJA) for wheat and May–August (MJJA) for maize and soybeans. See supplementary data for details.
[12] The multi-model econometric specification generates multi-model average responses, $\overline{\boldsymbol{\beta}}^T$ and $\overline{\boldsymbol{\beta}}^P$, controlling for variation among GGCMs via a model-specific indicator, γ: $y_{i,t,g} = \mu_i + \gamma_g + f(t) + \Sigma_j \overline{\beta}_j^T \xi_{j,i,t}^T + \Sigma_k \overline{\beta}_k^P \xi_{k,i,t}^P + \varepsilon_{i,t,g}$.
[13] The specification estimated using USDA data uses a state-specific time trend. While the ISIMIP-FT protocol requires management practices and technology to be held constant at year 2000 levels, different GGCMs include a variety of endogenous adaptation mechanisms (see section 3.5). We therefore consider a model with time effects more appropriate. For comparability, we also tested a specification for GGCMs using state-specific time trends as opposed to time effects (results available upon request). Results hold across different specifications.

between observed yields and meteorology within each interval, as well as the distribution of temperature and precipitation exposures across intervals in our transformed datasets.

Empirically-derived yield responses from the GGCM-ESM and USDA-PRISM datasets are not directly comparable because they are based on different meteorological inputs with distinct exposure distributions: ESM-simulated $\xi_j^{T,\mathrm{ESM}}$ and $\xi_k^{P,\mathrm{ESM}}$ versus observed $\xi_j^{T,\mathrm{PRISM}}$ and $\xi_k^{P,\mathrm{PRISM}}$. This raises the question of whether differences between the fitted GGCM-ESM and USDA-PRISM semi-elasticities ($\hat{\boldsymbol{\beta}}_g^T - \hat{\boldsymbol{\beta}}_{\mathrm{USDA}}^T$ and $\hat{\boldsymbol{\beta}}_g^P - \hat{\boldsymbol{\beta}}_{\mathrm{USDA}}^P$) are simply the product of differences in the distributions of temperature and precipitation inputs to yields (Q.IV). From equation (1), the weather-responsive component of log yield is defined as:

$$\psi_m(\boldsymbol{T}_i, \boldsymbol{P}_i) = \Sigma_j \hat{\beta}_{j,m}^T \xi_{j,i}^T + \Sigma_k \hat{\beta}_{k,m}^P \xi_{k,i}^P \tag{2}$$

and the difference between the weather-responsive components of GGCM and USDA yield is thus

$$\begin{aligned}\Delta\psi_g &= \psi_g(\boldsymbol{T}_i, \boldsymbol{P}_i) - \psi_{\mathrm{USDA}}(\boldsymbol{T}_i, \boldsymbol{P}_i)\\ &= \Sigma_j \hat{\beta}_{j,g}^T \xi_{j,i}^{T,\mathrm{ESM}} + \Sigma_k \hat{\beta}_{k,g}^P \xi_{k,i}^{P,\mathrm{ESM}}\\ &\quad - \left(\Sigma_j \hat{\beta}_{j,\mathrm{USDA}}^T \xi_{j,i}^{T,\mathrm{PRISM}} + \Sigma_k \hat{\beta}_{k,\mathrm{USDA}}^P \xi_{k,i}^{P,\mathrm{PRISM}}\right)\end{aligned} \tag{3}$$

Adding and subtracting cross-terms on the right-hand side of equation (3) and evaluating the weather exposure covariates at their 1981–2004 climatic means facilitates decomposition of $\Delta\psi$ into two terms, one capturing the effect of differences in climate forcing and the other capturing the effect of differing responses to meteorology:

$$\begin{aligned}\Delta\psi_g &= \underbrace{\Sigma_j \hat{\beta}_{j,g}^T \left(\xi_{j,i}^{T,\mathrm{ESM}} - \xi_{j,i}^{T,\mathrm{PRISM}}\right) + \Sigma_k \hat{\beta}_{k,g}^P \left(\xi_{k,i}^{P,\mathrm{ESM}} - \xi_{k,i}^{P,\mathrm{PRISM}}\right)}_{\text{Climate component } (\Delta\psi^{\mathrm{Climate}})}\\ &\quad \underbrace{+\Sigma_j \left(\hat{\beta}_{j,g}^T - \hat{\beta}_{j,\mathrm{USDA}}^T\right) \xi_{j,i}^{T,\mathrm{PRISM}} + \Sigma_k \left(\hat{\beta}_{k,g}^P - \hat{\beta}_{k,\mathrm{USDA}}^P\right) \xi_{k,i}^{P,\mathrm{PRISM}}}_{\text{Response component } (\Delta\psi^{\mathrm{Response}})}\end{aligned} \tag{4}$$

The relative importance of $\Delta\psi^{\mathrm{Climate}}$ and $\Delta\psi^{\mathrm{Response}}$ can then be assessed by comparing their distributions across locations.

Equation (1)'s estimated parameters enable us to investigate another key question: how do the characteristics of models drive the divergence between GGCM yield responses and those of historical yields to observed weather (Q.V). Drawing on documentation for each of our six GGCMs (Rosenzweig *et al* 2014, Elliott *et al* 2015), we construct binary indicator variables for five sets of characteristics likely to affect the yield response: (i) type of yield simulated (actual versus potential); (ii) endogenous cultivar change; (iii) heat stress; (iv) endogenous sowing date; (v) and whether the model was calibrated using site-specific or FAO country observations (supplementary table S6). We assemble characteristics (i)–(v) into a matrix, $\boldsymbol{Z}$. Then, using the stacked vector of temperature and precipitation semi-elasticities ($\zeta_m = [\hat{\boldsymbol{\beta}}_m^T, \hat{\boldsymbol{\beta}}_m^P]$) we compute the difference in the response from the USDA benchmark, $\Delta\zeta_g = \zeta_g - \zeta_{\mathrm{USDA}}$, which we employ as the dependent variable in the meta-analysis regression:[14]

$$\Delta\zeta = \boldsymbol{Z\eta} + \nu \tag{5}$$

The estimated parameters, η, indicate how strongly the shift in GGCM-ESM responses relative to the USDA-PRISM response is associated with each model attribute.

Finally, the implications of our estimated responses for future climate change impacts (Q.VI) are indicated by the yield changes that result from forcing our fitted empirical response functions with the distributions of temperature and precipitation under future climate warming. Log yield response functions from equation (2) are combined with meteorological exposures from bias-corrected HadGEM2-ES model simulations for our hindcast period (current climate), as well as mid-21st century (2033–2065) and late century (2067–2099) future climate under the RCP 8.5 (Moss *et al* 2010) high-warming scenario. In each epoch HadGEM2-ES daily temperature and precipitation ($\tilde{\boldsymbol{T}}_i$ and $\tilde{\boldsymbol{P}}_i$) fields are binned into the j and k intervals, respectively, to construct analogues of the weather exposure covariates, $\tilde{\xi}^T$ and $\tilde{\xi}^P$, for current and future years. Because climate simulations do not reproduce observed high-frequency weather extremes, and may exhibit biases relative to current climate (Vavrus *et al* 2015, Schoof and Robeson 2016), we do not directly compare simulated future exposures against their observed counterparts, but instead employ the 'delta' change

[14] This model is estimated with no constant. We test additional specifications to investigate both the impacts of model characteristics on the differences in responses to temperature and precipitation alone ($\zeta_m = \hat{\boldsymbol{\beta}}_m^T$ and $\zeta_m = \hat{\boldsymbol{\beta}}_m^P$, respectively), as well as the effects of interactions between characteristics and indicators of extreme high temperature and low precipitation.

method of computing differences in exposure between ESM-simulated current and future climates[15]. Specifically, we time-average the temperature and precipitation bins to generate the mean meteorological exposure for the hindcast period (current climate), calculate the difference between the resulting average and simulated exposure under future climate, and finally multiply the result by the estimated semi-elasticities to generate meteorological shocks to log yields. We use the latter to compute a normalized multi-decadal index of climate impact, given by the ratio of each location's average yield under a future climate to its average yield under the present climate. Using $\mathbb{E}$ to denote the expected value over each epoch, the index is:

$$\Psi_{i,m} = \mathbb{E}\left[\exp\left\{\psi_m\left(\tilde{\boldsymbol{T}}_i^{\text{Future Climate}}, \tilde{\boldsymbol{P}}_i^{\text{Future Climate}}\right) - \psi_m\left(\tilde{\boldsymbol{T}}_i^{\text{Current Climate}}, \tilde{\boldsymbol{P}}_i^{\text{Current Climate}}\right)\right\}\right] \quad (6)$$

We note that Ψ_i diverges from fractional changes in future yields from the current climate projected by GGCMs, as equation (6) omits both the CFE and endogenous adaptation mechanisms into GGCMs models, particularly endogenous or unrecorded prescribed future changes in fertilizer application rates, crop calendars, or crop genotypes[16].

3. Results

3.1. GGCMs' ability to reproduce recorded yields

Figure 1 summarizes the distributions of the differences in percentage yield anomalies between GGCMs and USDA records for our three crops over the 1981–2004 period. The wide support of the distribution suggests that the ISIMIP-FT GGCMs struggle to reproduce the PDF of historical US yield anomalies. For counties within the interquartile range the GGCM-observation divergence is ±30%, while in the majority of remaining locations simulated yields can dramatically overstate or understate the observations.

While this pattern persists across crops, GGCMs' performance—as indicated by the variance of the distributions—is generally better for wheat and especially maize compared to soybeans. The modes of the individual annual cross-county PDFs (shown in light colors) exhibit positive and negative interannual fluctuations, but do not follow any easily discernible pattern that suggests systematic bias. The differences across models and among crops in the annual and aggregate PDFs also suggest that no single GGCM has a clear advantage in modeling all crops[17]. A certain GGCM may exhibit skill in modeling a particular crop (e.g. LPJmL wheat), while some GGCMs outperform others in simulating a certain crop (e.g. GAEZ-IMAGE versus GEPIC for maize).

3.2. Yield correlations with adverse weather extremes: simulations vs. observations

A more nuanced way to evaluate GGCMs' performance is to examine how well they reproduce historical correlations between annual yield anomalies and exposure to extreme high temperature and low precipitation. We do this in figure 2 by presenting the correlations between de-trended yields and annual growing season exposures to extreme high temperature and extreme low precipitation bins as a bivariate PDF. Relative to our comparison of yield anomalies (section 3.1), there is more agreement in correlations between ESM-simulated meteorological extremes and GGCM-simulated yields, and the correlations between PRISM meteorological extremes and observed yields. Both correlations are negative in 50%–75% of counties (with the exception of GAEZ-IMAGE), and the magnitudes of the correlations differ both across models and among crops. Simulated maize and soybean responses are for the most part qualitatively similar to observations, with GEPIC, LPJ-GUESS, LPJmL showing tight clustering of negative correlations across counties. Even so, simulated wheat responses vary markedly relative to one another, and diverge from observations. This result may arise from GGCMs simulating different types of wheat (e.g. GGCMs decide internally the type of wheat to be grown) while our observational data are spring durum wheat only.

3.3. Simulated and observed yield responses to weather

In a refinement of the analysis in section 3.2 we statistically model additional factors that affect yield. One is management practices, whose sub-national and interannual variation is unfortunately not available in either the GGCM-ESM or USDA-PRISM datasets. Another is non-extreme weather: negative yield impacts of more frequent extreme low precipitation and/or high temperature days might be offset by

[15] First studies using this method include Arnell (1996) and Gleick (1986). For application of this method in the context of agriculture see (Roberts *et al* 2013).

[16] For instance, see Rosenzweig *et al* (2014) SI for details on adaptations accounted for by the GGCMs, and Elliott *et al* (2015) for revised protocols in the next phase of GGCMs' simulations to introduce harmonization in GGCMs' simulation runs.

[17] GAEZ-IMAGE appears to be an exception, perhaps due to its unique temporal scale relative to other GGCMs—interpolating monthly meteorology to a daily time-step, while simulating annual yields every 5th year and interpolating yields for the intervening years (Rosenzweig *et al* 2014: table S4).

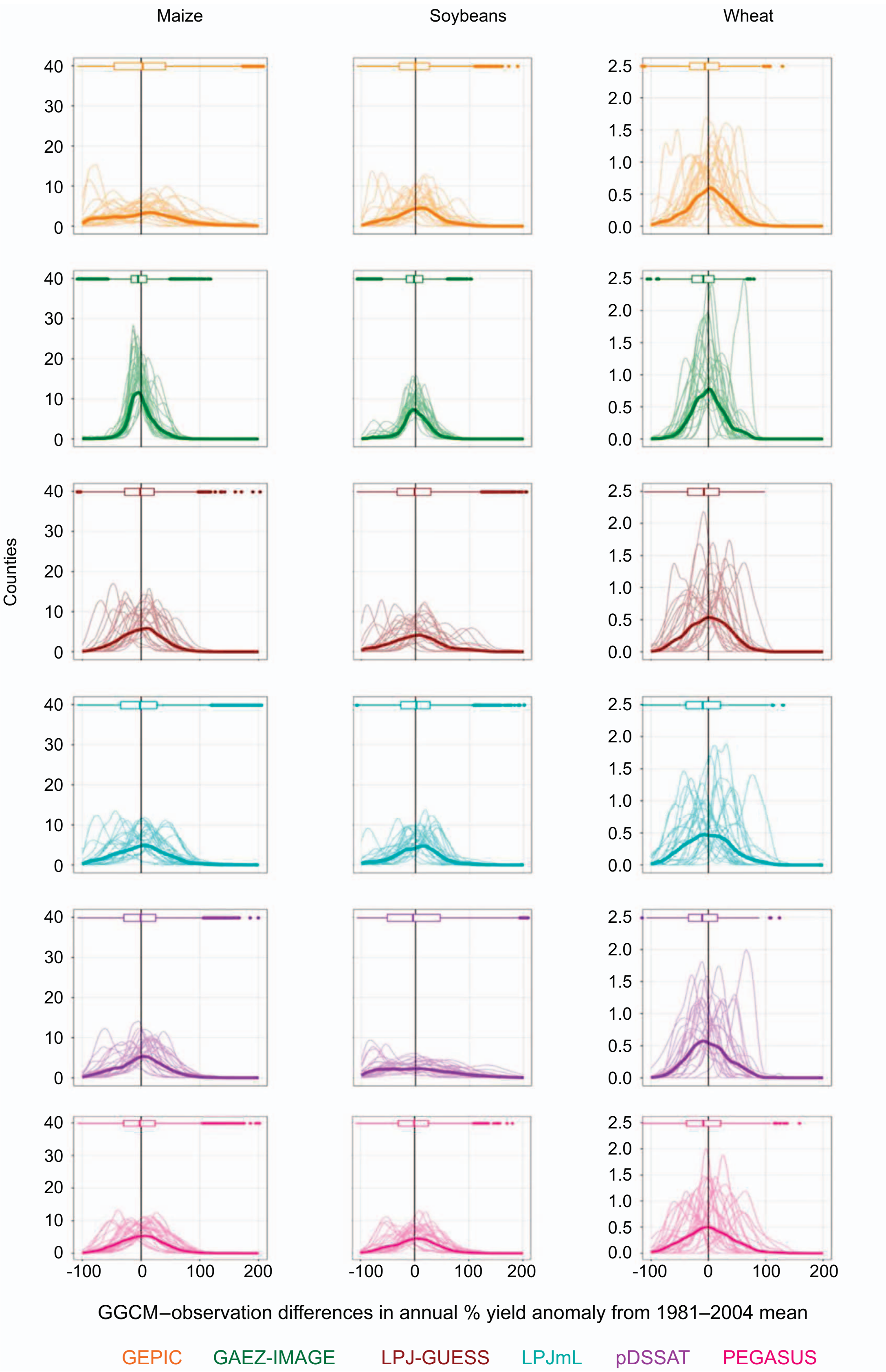

Figure 1. Cross-county distribution of the GGCM–USDA difference in percentage yield anomalies for maize, soybeans, and wheat. Anomalies are calculated as the % deviation of each county's de-trended yield from its own 1981–2004 mean. Light lines show the annual distribution of county differences between each model and observations. Heavy lines show the ditribution across counties and years.

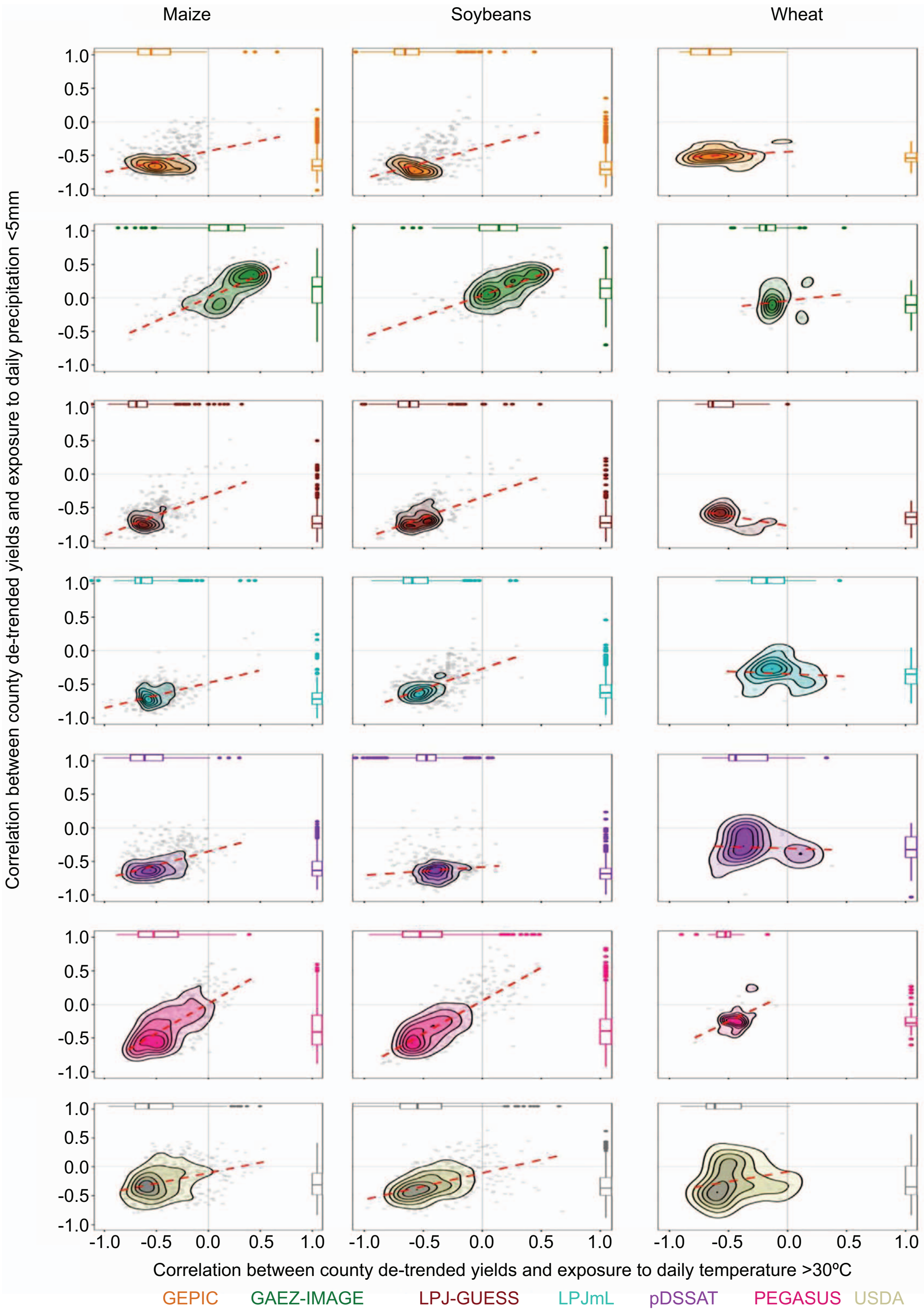

Figure 2. Correlations between de-trended yields and extreme high temperature exposures (horizontal axis) and extreme low precipitation exposures (vertical axis) for six GGCMs and observations. Dashed red lines are the linear fit indicating the cross-county pattern of association between temperature and precipitation exposure correlations.

near-optimal growing conditions throughout the remainder of the growing season, while yields may be lower in counties and years that experience fewer extreme adverse days, but more frequent non-extreme but nonetheless sub-optimal weather.

Equation (1) accounts for both sets of factors by partitioning the variance in yields into influences associated with unobservables (μ_i and $f(t)$) and the mean deterministic effects of the distribution of temperature and precipitation conditions experienced by crops. Figure 3 illustrates the splines tracing out the responses of log yield to the distribution of temperature and precipitation. All covariates explain 75% of the cross-section/time-series yield variation

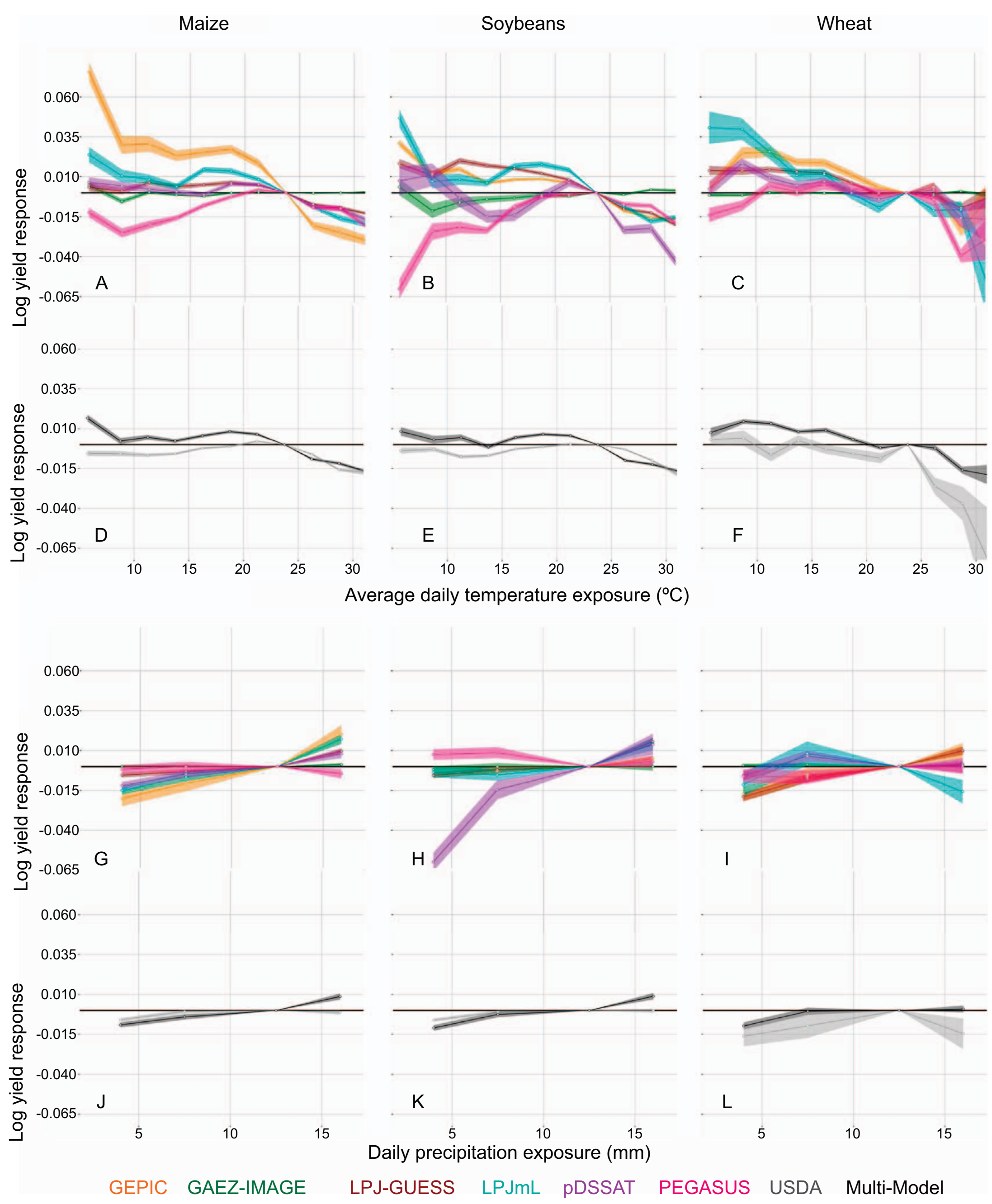

Figure 3. Mean responses (solid lines) and confidence intervals (95%) (shaded areas) of log yield to temperature and precipitation exposure for maize, soybeans and wheat (equation (1)). Responses are normalized relative to the number of days with temperatures 22.5 °C–25 °C and precipitation 10–15 mm day^{-1}, represented by the heavy horizontal axis. Shaded confidence intervals are computed from robust standard errors clustered at cross-sectional units (grid-cell for GGCMs, counties for USDA).

(supplementary table S4), and the weather responses account for between 0% and 60% (supplementary table S5). GGCM and USDA yield responses are both consistent with empirical findings on the negative effects of exposure to high daily temperatures and (aside from GEPIC maize and pDSSAT soybean simulations) as well as smaller magnitude responses to low precipitation (Schlenker and Roberts 2009, Tack *et al* 2015).

Whether the responses of different GGCMs to both extreme and non-extreme weather vary can be said to diverge from one another (panels A–C and G–I), and from the USDA-PRISM benchmark (panels D–F and J–L) depends on the specification of the variance-covariance matrix of the error term in equation (1). Our default standard errors are clustered at the level of cross-sectional units (counties in the case of USDA-PRISM and grid-cells in the case of GGCMs) and are robust to temporal autocorrelation. They suggest differences in responses among individual GGCMs, and between GGCMs and USDA-PRISM that are statistically significant (supplementary table S8). However, in empirical models of crop yields, residual spatial autocorrelation can substantially inflate the standard errors of the coefficients (Yun *et al* 2015). Adjusting for joint residual temporal and spatial autocorrelation using Cameron *et al* (2011) clustering of the standard errors by county/cell and year increases their values by factors of 2–3 (supplementary table S4), weakening the conclusion that the GGCM and USDA-PRISM responses significantly diverge—especially in the case of extreme high-temperatures (Schauberger *et al* 2017), but less so for extreme low precipitation (supplementary table S8). Even so, for either specification of the variance-covariance matrix, no GGCM exhibits a consistent positive or negative bias relative to the USDA-PRISM response.

The USDA-PRISM response suggests that exposure to an additional day >30 °C reduces annual maize and soybean yields by 1.5% but generates wheat yield losses six times as large. For GGCMs, the corresponding response varies between 0.2%–3% for maize, 0.5%–3.6% for soybeans, and 0.1%–6.5% for wheat, and the observed responses fall within the range of simulated responses, except for wheat. Exposure to an additional day with precipitation <5 mm reduces maize and soybean yields by about 0.5% and wheat by about 1.5% in the observational dataset. GGCMs exhibit larger losses for maize and soybeans (with the exception of PEGASUS), between 0 and 4.5% (1% at the multi-model average response), whereas wheat's response to dry days in the observational dataset is understated by most models (with the exception of GEPIC and LPJmL)[18].

3.4. Decomposition of the divergence between GGCM and USDA yield responses

We focus on two factors that likely drive the GGCM-observation divergence in figure 3.[19] The first is differences between the aggregate responses to weather shocks implied by process models' internal representation of crop growth and the responses of observed agricultural systems. The second is differences in the exposures implied by the PRISM data for the observations as opposed to HadGEM2-ES for the GGCMs. We use the decomposition technique illustrated in equation (4) to establish their relative magnitudes. Figure 4 shows the results of this calculation.

The horizontal axis rank-orders counties from the largest negative to the largest positive values of the difference between the weather-responsive portion of each GGCM's historical run and the observations, $\Delta\psi$, whose magnitude is measured on the vertical axis and whose county values are indicated by black dots. For each county the corresponding light- and dark-colored bars indicate the response and climatic components of the divergence ($\Delta\psi^{\mathrm{Response}}$ and $\Delta\psi^{\mathrm{Climate}}$, respectively). For the majority of GGCM x crop combinations, cross-county trends in the total divergence and $\Delta\psi^{\mathrm{Response}}$ closely track one another, while $\Delta\psi^{\mathrm{Climate}}$ tends to add either noise or an offset. This result demonstrates that the differences in the splines in figure 3 are mostly attributable to GGCMs' internal responses, not differences in meteorological inputs.

3.5. Correlates of the GGCM-USDA yield response divergence

Finally, table 1 summarizes our meta-analytic results that associate model attributes with the gaps between GGCMs' responses and those derived from historical observations. To conserve space we report results for maize only, and consign results for wheat and soybeans in the supplementary data (table S7). The largest magnitude coefficient is on heat stress, whose overall impact is to make the divergence in responses the more negative, suggesting that in panels A and G of figure 3, the responses of the sole model incorporating this mechanism (PEGASUS) exhibits a smaller change in yield (i.e. a downward shift) for an additional day of exposure over their entire range of weather variation. Simultaneously, the positive effect of heat stress interacted with high-temperature (low-precipitation) intervals indicates that in panel A (G) the right (left) tails of the corresponding splines are shifted upward, resulting in a less weather sensitive—i.e. flatter—response profile. Cultivar adaptation, the second largest influence, acts in the opposite way: inducing an upward shift in the response profiles over their entire range that is outweighed by the negative impact of interactions with extreme high temperature and low precipitation exposures, resulting in a more weather sensitive—i.e. steeper—profile for models that include this mechanism (GEPIC, and less evident for LPJ-GUESS, PEGASUS). Other characteristics, such as endogenous selection of sowing dates and model calibration based on site-specific studies—which respectively flatten and steepen the response profiles, have a smaller overall influence and are not uniformly significant across all crops. The major implication is that with a flatter response profile, shifts in the distributions of temperature and precipitation inputs translate into smaller simulated yield changes, while a steeper response profile can result in excess sensitivity that translates modest weather shocks into large yield changes.

We obtain broadly similar results for soybeans, but equivocal estimates for wheat (supplementary table S7), whose response is positively affected by heat stress interacted with low precipitation intervals, capturing the PEGASUS model's flatter response to precipitation relative to the other GGCMs.

[18] The econometric models for simulated wheat generally have a lower explanatory power compared to maize and soybeans (see table S5). This might be due to differences in the type of wheat chosen by models compared to the variety observed, spring durum wheat) and to the fact that those varieties might be grown outside the growing season (April–August), see also section S5 in the supplementary data.

[19] A potential third issue is omitted variable bias, in the form of contaminating effects on the estimated parameters of management practices that are correlated with weather and unrecorded in the observational dataset, but omitted from GGCM simulations.

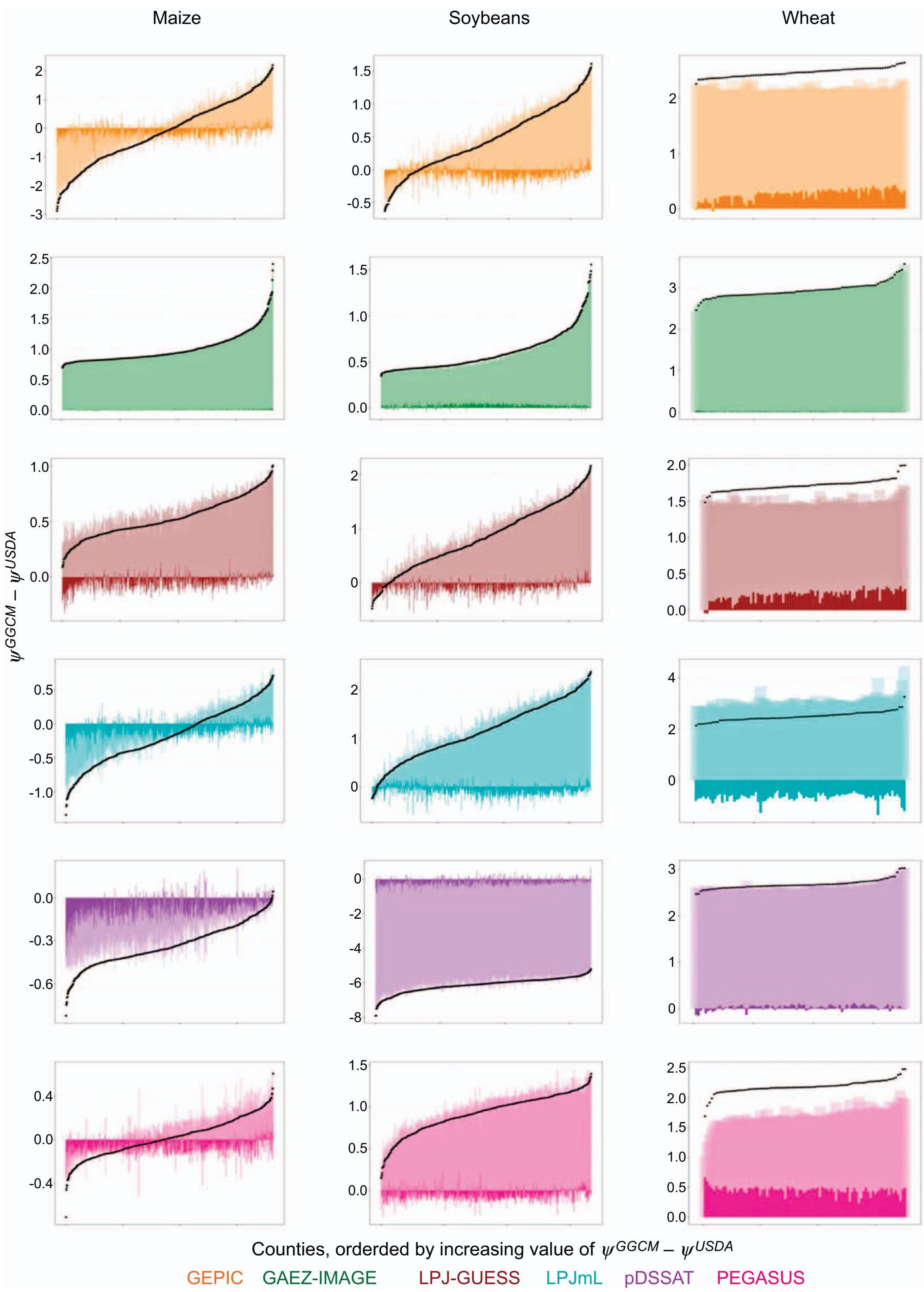

Figure 4. Decomposition of predicted weather component of GGCM yield-predicted weather component of observed yield for counties (about 730 for maize, 670 for soybeans, 88 for wheat) showing the total difference (black line), climate component (dark bars), and response component (light bars).

While the source of this disparity is not clear cut, we speculate that it emanates from inter-model variation in the type of wheat being grown, and emphasize that our meta-analytic approach will likely prove more beneficial in imminent intercomparison exercises with comprehensive records (e.g. ISI-MIP2 and Global Gridded Crop Model Intercomparison, Elliott *et al* 2015).

4. Discussion and conclusions

Using cross-section/time-series datasets of simulated and observed rainfed yields of maize, wheat and soybeans for about 1000 US counties over 24 years, we have characterized the heterogeneous responses of crop models to ESM-simulated temperature and precipitation, and compared them with empirically

Table 1. Effects of model characteristics on GGCM-USDA divergence in maize yield response. Model specifications are discussed in the supplementary data. Robust standard errors in parentheses. Table S7 summarizes results for soybeans and wheat.

Dependent variable	$\Delta\zeta[\hat{\beta}^P, \hat{\beta}^T]$	$\Delta\zeta[\hat{\beta}^T]$	$\Delta\zeta[\hat{\beta}^P]$	$\Delta\zeta[\hat{\beta}^P, \hat{\beta}^T]$	$\Delta\zeta[\hat{\beta}^T]$	$\Delta\zeta[\hat{\beta}^P]$
Potential yield	−0.007	−0.011	0.004	−0.024***	−0.028***	−0.015***
	(0.008)	(0.009)	(0.010)	(0.008)	(0.008)	(0.004)
Endog. cultivar	0.013	0.017*	−0.001	0.033***	0.036*	0.021
	(0.008)	(0.010)	(0.010)	(0.008)	(0.008)	
Endog. sowing date	−0.0003	−0.0004		−0.005***	−0.005**	
	(0.002)	(0.002)		(0.002)	(0.002)	
Heat stress	−0.017**	−0.023**	0.001	−0.035***	−0.040***	−0.025
	(0.008)	(0.009)	(0.010)	(0.008)	(0.008)	
Site calibration	0.004**	0.005***		0.007***	0.005***	
	(0.002)	(0.001)		(0.001)	(0.002)	
T >30 °C						
× Potential yield				0.051***	0.059***	
				(0.009)	(0.009)	
× Endog. cultivar				−0.056***	−0.063***	
				(0.009)	(0.009)	
× Endog. sowing date				0.015***	0.015***	
				(0.004)	(0.004)	
× Heat stress				0.48***	0.56***	
				(0.009)	(0.009)	
× Site calibration				−0.006**		
				(0.003)		
P < 5 mm						
× Potential yield				0.035***		0.028***
				(0.008)		(0.005)
× Endog. cultivar				−0.044***		−0.033***
				(0.008)		(0.002)
× Endog. sowing date				0.010***		
				(0.003)		
× Heat stress				0.44***		0.039***
				(0.008)		(0.002)
				−0.012***		
				(0.002)		
F Adj.	4.289***	4.885***	1.553	42.577***	24.052***	2.325*
	(df = 4;77)	(df = 4;59)	(df = 2;17)	(df = 14;77)	(df = 8;59)	(df = 5;17)
Obs.	78	60	18	78	60	18
Adj. R Sq.	0.211	0.308	−0.153	0.593	0.624	0.336

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

derived responses to observed weather series. The six GGCM simulations we examined do not reproduce the cross-county, inter-annual distributions of yield. Notwithstanding this, our econometric analyses indicate that GGCM broadly capture the major stylized facts of weather impacts on crop yields that have been identified by the empirical climate change economics literature. Yet the responses of individual GGCMs differ substantially from one another and relative to their observationally-derived counterpart. Simulated yields are generally more temperature sensitive than observed yields, but can be more or less sensitive to high temperature or low precipitation extremes, depending on the particular model and crop. We show that such behavior is attributable to differences in how models simulate heat stress and cultivar adaptation. GGCMs incorporating the latter (former) mechanism tend to be more (less) sensitive to weather shocks.

The consequences of these details for the impacts of climate change on US crops are summarized in figure 5. The yield changes therein are calculated not by running GGCMs with meteorological inputs projected by ESMs, but by forcing their response functions derived in figure 3 with changes in future temperature and precipitation exposures from the historical period simulated by HadGEM2-ES. They therefore do not account for the potential benefits of the CFE, or future management changes and other adaptations either endogenously computed by, or exogenously imposed upon GGCMs simulations as part of the ISIMIP-FT exercise. Notwithstanding the overlap in the confidence intervals of the GGCMs' responses, under vigorous warming, late-century (2067∼2099) projections of production changes based on the coefficient point estimates diverge widely; ranging from −96% to +6%—and −71% at the multi-model mean response—for maize,

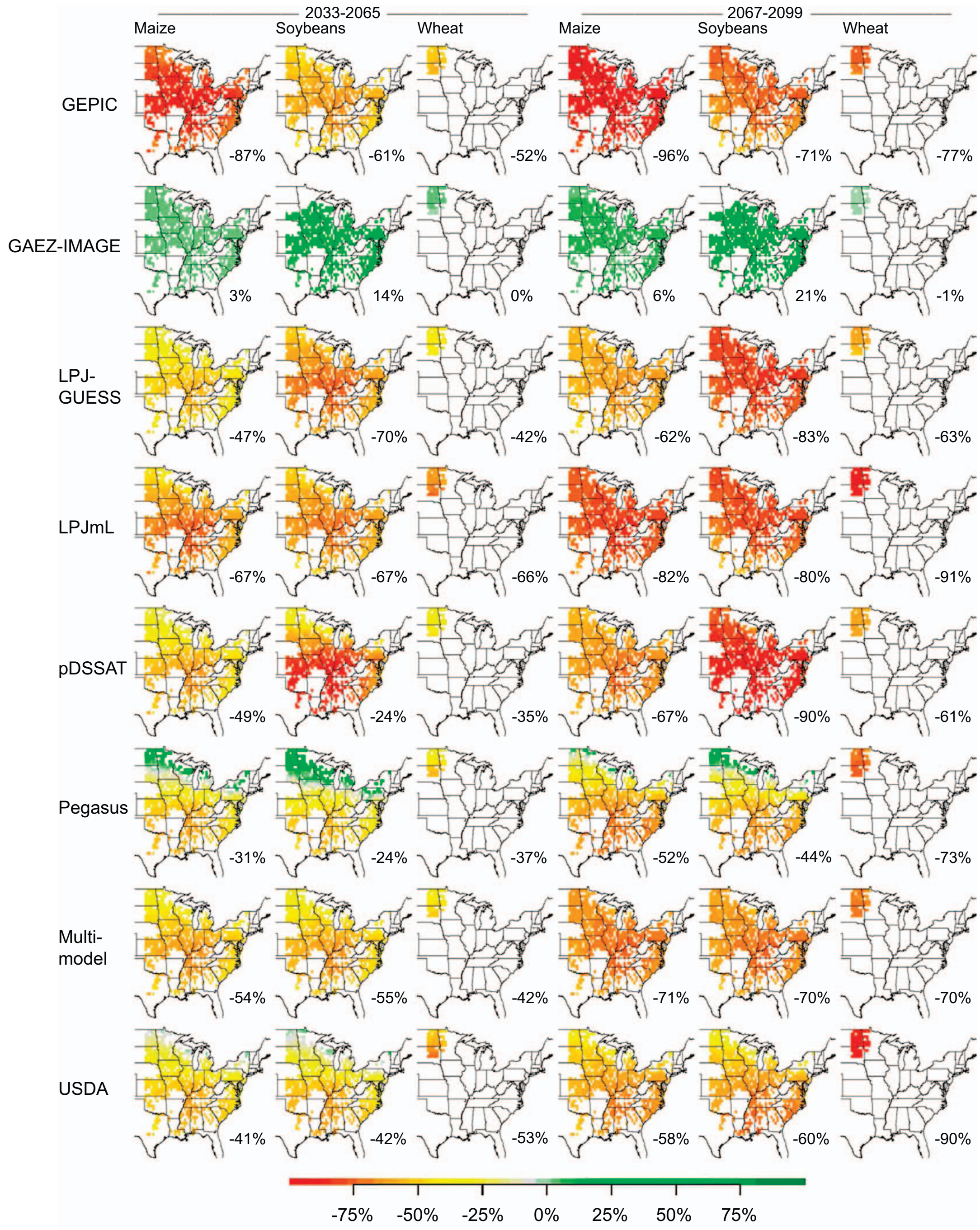

Figure 5. Mid- and end-century % change in rainfed yields under RCP 8.5 warming scenario simulated by HadGEM2-ES. The % change numbers accompanying each map depict the projected % change in aggregated production across the sample of counties, under the assumption of same harvested area in future periods, as in historical.

−90% to +21% with a mean of −70% for soybeans, −91% to −1% with a mean of −70% for wheat. The responses of GGCMs that are most sensitive to extreme high temperatures (GEPIC, LPJ-GUESS and LPJmL) are associated with the largest losses, in excess of 40% of maize and wheat production, and 60% of soybean production by mid-century (2033∼2065), while only GAEZ-IMAGE predicts production gains. Relative to the GGCM responses, our USDA-PRISM response generates smaller losses (−58% for maize, −60% for soybeans, −90% for wheat) for late-century, but its predicted production declines due to more frequent days > 30°C closely track those reported by Schauberger *et al* (2017)[20] for the 30°C – 36°C temperature range (−54% for maize, −60% for soybeans and −73% for wheat—see supplementary table S9) which gives us confidence in the reliability of our approach[21].

[20] While a direct comparison with results of Schauberger *et al* (2017) is difficult to make for maize and soybeans (due to a larger number of counties utilized in their study), results for wheat are not comparable due to winter wheat used in their study.

[21] By contrast, GGCMs' late century (2067∼2099) losses due to extreme high temperature days (> 30 °C), range from −72% to +3%—and −53% at the multi-model mean—for maize, −86% to +7% with a mean of −56% for soybeans, and −66% to −4% with a mean of −42% for wheat.

By relying solely on meteorological inputs, and ignoring confounding factors such as the CFE, exogenous future adaptations or additional endogenous adjustments such as shifts in cultivars and crop calendars represented within models, our projections provide insights into how GGCMs' characteristics can amplify or moderate climatically-driven yield declines. For example, a key feature of figure 5 is the lack of spatial (particularly latitudinal) variation in GGCM yield shocks compared to the USDA projections. The exception is the PEGASUS model, whose flatter response profiles generate smaller losses than the USDA benchmark. For most of the remaining GGCM responses the converse is true: excess sensitivity generates yield changes—and, without compensating adaptation mechanisms, production losses—that are uniformly large. Heat stress at anthesis (and, secondarily, endogenous sowing) may therefore be important for bringing models' overall sensitivity into better agreement with the responses exhibited by observed agricultural systems. But this also raises the question of what model attributes might drive GGCMs' excess sensitivity. Our findings hint at endogenous cultivar selection as a potential candidate, as it amplifies negative yield responses to low precipitation in soybeans, high temperature in wheat, and both types of weather shocks in maize. Another may be the use of site-specific data for calibrating maize and soybean simulations, but the potential mechanisms are unclear.

Such interpretation challenges highlight four important caveats to our analysis. The first is the small number of observations on which our meta-analysis results are based, especially relative to the number of dimensions along which GGCMs can potentially vary. Without a larger sample of models, little can be done to increase the statistical power of our assessment. A second, related issue is that because the ISIMIP-FT protocol did not mandate standardization of GGCMs' characteristics, or harmonization and recording of the corresponding detailed inputs across models and scenarios, our own coding of model attributes could conceivably introduce errors. Third, the aforementioned paucity of data required us to use all of the parameters of the GGCM and USDA-PRISM estimated responses, as opposed to zeroing out differences that were not statistically significant. With the latter approach, the substantial reduction in the divergence between GGCM- and observationally-based responses when residual spatial autocorrelation is accounted for can potentially weaken our inferences in table 1. Finally, because the GGCM simulations employed here are not specifically optimized for US counties, it is not clear how well our results extrapolate beyond the specific spatial domain of the eastern US.

All of these limitations are already being addressed by the current generation of crop model intercomparison exercises (ISI-MIP2, the Global Gridded Crop Model Intercomparison (Elliott *et al* 2015)), which are in the process of fielding larger numbers of GGCMs running more controlled experiments with considerable efforts being made to harmonize and record key inputs such as management practices, and evaluate model outputs against a common set of recently-developed global historical data benchmarks (Ray *et al* 2012, Iizumi *et al* 2014). Our hope is that the inter-method comparison techniques developed here can contribute to improving the evaluation of the results of these exercises (cf Müller *et al* 2017), with the goals of more rigorously pinpointing the origins of GGCMs' emergent crop yield responses, and thereby strengthening the empirical basis of global-scale assessment of future climate change impacts on agriculture.

Acknowledgments

We thank the ISI-MIP Fast Track Agriculture modeling teams, Christoph Müller and Joshua Elliott for their assistance with the data. Participants in focus issue workshops at the European Union Joint Research Centre Seville and Stanford University, and three anonymous referees provided helpful comments and suggestions. Ian Sue Wing was supported by the US National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507, and by US Department of Energy Office of Science (BER) Integrated Assessment Research Program, grants DE-SC0005171 and DE-SC0016162.

ORCID

Malcolm N Mistry https://orcid.org/0000-0003-3345-6197
Enrica De Cian https://orcid.org/0000-0001-7134-2540

References

Arnell N 1996 *Global Warming, River Flows and Water Resources* (Chichester: Wiley)

Bassu S *et al* 2014 How do various maize crop models vary in their responses to climate change factors? *Glob. Change Biol.* **20** 2301–20

Blanc É 2017 Statistical emulators of maize, rice, soybean and wheat yields from global gridded crop models *Agric. Forest Meteorol.* **236** 145–61

Blanc E and Sultan B 2015 Agricultural and forest meteorology emulating maize yields from global gridded crop models using statistical estimates *Agric. Forest Meteorol.* **214–5** 134–47

Bondeau A *et al* 2007 Modelling the role of agriculture for the 20th century global terrestrial carbon balance *Glob. Change Biol.* **13** 679–706

Bouwman A F, Kram T and Klein Goldewijk K 2006 Integrated modelling of global environment change. An overview of IMAGE 2.4

Burke M and Emerick K 2016 Adaptation to climate change: evidence from US agriculture *Am. Econ. J. Econ. Policy* **8** 106–40

Cameron C, Gelbach J and Miller D 2011 Robust inference with multiway clustering *J. Bus. Econ. Stat.* **29** 238–49

Deryng D, Sacks W J J, Barford C C C and Ramankutty N 2011 Simulating the effects of climate and agricultural management practices on global crop yield *Glob. Biogeochem. Cycles* **25** GB2006

Deschênes O and Greenstone M 2007 The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather *Am. Econ. Rev.* **97** 354–85

Deschênes O and Greenstone M 2012 The economic impacts of climate change: evidence from in weather output and random fluctuations agricultural *Am. Econ. Rev.* **97** 354–85

Elliott J, Kelly D, Best N, Wilde M, Glotter M and Foster I 2013 The parallel system for integrating impact models and sectors (pSIMS) *Proc. of the Conf. on Extreme Science and Engineering Discovery Environment: Gateway to Discovery XSEDE '13* (New York, NY: ACM) pp 21:1–21:8

Elliott J *et al* 2015 The global gridded crop model intercomparison: data and modeling protocols for phase 1 (v1.0) *Geosci. Model Dev.* **8** 261–77

Frieler K *et al* 2015 A framework for the cross-sectoral integration of multi-model impact projections: land use decisions under climate impacts uncertainties *Earth Syst. Dyn.* **6** 447–60

Gleick P H 1986 Methods for evaluating the regional hydrologic impacts of global climatic changes *J. Hydrol.* **88** 97–116

Gornall J, Betts R, Burke E, Clark R, Camp J, Willett K and Wiltshire A 2010 Implications of climate change for agricultural productivity in the early twenty-first century *Philos. Trans. R. Soc. London* B **365** 2973–89

Hempel S, Frieler K, Warszawski L, Schewe J and Piontek F 2013 A trend-preserving bias correction—the ISI-MIP approach *Earth Syst. Dyn.* **4** 219–36

Iglesias A, Rosenzweig C and Pereira D 2000 Agricultural impacts of climate change in Spain: developing tools for a spatial analysis *Glob. Environ. Change* **10** 69–80

Iizumi T, Yokozawa M, Sakurai G, Travasso M I, Romanenkov V, Oettli P, Newby T, Ishigooka Y and Furuya J 2014 Historical changes in global yields: major cereal and legume crops from 1982 to 2006 *Glob. Ecol. Biogeogr.* **23** 346–57

IPCC 2014 *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* ed C B Field, V R Barros, D J Dokken and K J Mach (Cambridge and New York, NY: Cambridge University Press)

Jones C D *et al* 2011 The HadGEM2-ES implementation of CMIP5 centennial simulations *Geosci. Model Dev.* **4** 543–70

Jones J W *et al* 2016 Brief history of agricultural systems modeling *Agr. Syst.* accepted (https://doi.org/10.1016/j.agsy.2016.05.014)

Jones J W, Hoogenboom G, Porter C H, Boote K J, Batchelor W D, Hunt L A, Wilkens P W, Singh U, Gijsman A J and Ritchie J T 2003 The {DSSAT} cropping system model *Eur. J. Agron.* **18** 235–65

Liu J, Williams J R, Zehnder A J B and Yang H 2007 GEPIC—modelling wheat yield and crop water productivity with high resolution on a global scale *Agric. Syst.* **94** 478–93

Lobell D B and Asseng S 2017 Comparing estimates of climate change impacts from process-based and statistical crop models *Environ. Res. Lett.* **12** 015001

Lobell D B and Burke M B 2010 On the use of statistical models to predict crop yield responses to climate change *Agric. Forest Meteorol.* accepted (https://doi.org/10.1016/j.agrformet.2010.07.008)

Lobell D B, Schlenker W and Costa-Roberts J 2011 Climate trends and global crop production since 1980 *Science* **333** 616–20

Lobell D B, Sibley A and Ivan Ortiz-Monasterio J 2012 Extreme heat effects on wheat senescence in India *Nat. Clim. Change* **2** 186–9

Moore F C and Lobell D B 2015 The fingerprint of climate trends on European crop yields *Proc. Natl Acad. Sci.* **112** 2670–5

Morell F J *et al* 2016 Can crop simulation models be used to predict local to regional maize yields and total production in the US Corn Belt? *Field Crop. Res.* **192** 1–12

Moss R H *et al* 2010 The next generation of scenarios for climate change research and assessment *Nature* **463** 747–56

Müller C *et al* 2017 Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications *Geosci. Model Dev.* **10** 1403–22

Müller C, Elliott J, Chryssanthacopoulos J, Deryng D, Folberth C, Pugh T A M and Schmid E 2015 Implications of climate mitigation for future agricultural production *Environ. Res. Lett.* **10** 125004

Ortiz-Bobea A 2013 Is weather really additive in agricultural production? Implications for climate change impacts *Resources For the Future Discussion Papers* (http://econpapers.repec.org/RePEc:rff:dpaper:dp-13-41) (https://doi.org/10.2139/ssrn.2374506)

Oyebamiji O K, Edwards N R, Holden P B, Garthwaite P H, Schaphoff S and Gerten D 2015 Emulating global climate change impacts on crop yields *Stat. Model.* **15** 499–525

Porter J R, Xie L, Challinor A J, Cochrane K, Howden S M, Iqbal M M, Lobell D B and Travasso M I 2014 Food security and food production systems *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge: Cambridge University Press) pp 485–533

Ray D K, Ramankutty N, Mueller N D, West P C and Foley J A 2012 Recent patterns of crop yield growth and stagnation *Nat. Commun.* **3** 1293

Roberts M J, Schlenker W and Eyer J 2013 Agronomic weather measures in econometric models of crop yield with implications for climate change *Am. J. Agric. Econ.* **95** 236

Rosenzweig C *et al* 2014 Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison *Proc. Natl Acad. Sci. USA* **111** 3268–73

Schauberger B *et al* 2017 Consistent negative response of US crops to high temperatures in observations and crop models *Nat. Commun.* **8**

Schlenker W and Lobell D B 2010 Robust negative impacts of climate change on African agriculture *Environ. Res. Lett.* **5** 014010

Schlenker W and Roberts M J 2006 Estimating the impact of climate change on crop yields: The importance of nonlinear temperature effects *NBER Working Paper* 13799 (www.nber.org/papers/w13799)

Schlenker W and Roberts M J 2009 Nonlinear temperature effects indicate severe damages to US crop yields under climate change *Proc. Natl Acad. Sci. USA* **106** 15594–8

Schoof J T and Robeson S M 2016 Projecting changes in regional temperature and precipitation extremes in the United States *Weather Clim. Extrem.* **11** 28–40

Sitch S *et al* 2003 Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model *Glob. Change Biol.* **9** 161–85

Tack J, Barkley A and Nalley L L 2015 Effect of warming temperatures on US wheat yields *Proc. Natl Acad. Sci.* **112** 6931–6

Vavrus S J, Notaro M and Lorenz D J 2015 Interpreting climate model projections of extreme weather events *Weather Clim. Extrem.* **10** 10–28

Warszawski L, Frieler K, Huber V, Piontek F, Serdeczny O and Schewe J 2013 The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework 1–5

Watson J, Challinor A J, Fricker T E and Ferro C A T 2015 Comparing the effects of calibration and climate errors on a statistical crop model and a process-based crop model *Clim. Change* **132** 93–109

Sue Wing I, Monier E, Stern A and Mundra A 2015 US major crops' uncertain climate change risks and greenhouse gas mitigation benefits *Environ. Res. Lett.* **10** 115002

Yun S, Gramig B, Delgado M and Florax R G J M 2015 Does spatial correlation matter in econometric models of crop yield response and weather? *Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting (San Francisco, CA)*