G Model
AGMET-4285; No. of Pages 10
ARTICLE IN PRESS
Agricultural and Forest Meteorology xxx (2010) xxx–xxx
Contents lists available at ScienceDirect
Agricultural and Forest Meteorology
journal homepage: www.elsevier.com/locate/agrformet

On the use of statistical models to predict crop yield responses to climate change

David B. Lobell[a,b,*], Marshall B. Burke[b,c]

[a] Department of Environmental Earth System Science, Stanford University, Stanford, CA 94305, United States
[b] Program on Food Security and Environment, Stanford University, Stanford, CA 94305, United States
[c] Department of Agricultural and Resource Economics, University of California, Berkeley, CA 94720, United States

ARTICLE INFO

Article history:
Received 19 October 2009
Received in revised form 16 June 2010
Accepted 20 July 2010

Keywords:
CERES-Maize
Africa
Marksim
Maize

ABSTRACT

Predicting the potential effects of climate change on crop yields requires a model of how crops respond to weather. As predictions from different models often disagree, understanding the sources of this divergence is central to building a more robust picture of climate change's likely impacts. A common approach is to use statistical models trained on historical yields and some simplified measurements of weather, such as growing season average temperature and precipitation. Although the general strengths and weaknesses of statistical models are widely understood, there has been little systematic evaluation of their performance relative to other methods. Here we use a perfect model approach to examine the ability of statistical models to predict yield responses to changes in mean temperature and precipitation, as simulated by a process-based crop model. The CERES-Maize model was first used to simulate historical maize yield variability at nearly 200 sites in Sub-Saharan Africa, as well as the impacts of hypothetical future scenarios of 2 °C warming and 20% precipitation reduction. Statistical models of three types (time series, panel, and cross-sectional models) were then trained on the simulated historical variability and used to predict the responses to the future climate changes. The agreement between the process-based and statistical models' predictions was then assessed as a measure of how well statistical models can capture crop responses to warming or precipitation changes. The performance of statistical models differed by climate variable and spatial scale, with time-series statistical models ably reproducing site-specific yield response to precipitation change, but performing less well for temperature responses. In contrast, statistical models that relied on information from multiple sites, namely panel and cross-sectional models, were better at predicting responses to temperature change than precipitation change. The models based on multiple sites were also much less sensitive to the length of historical period used for training. For all three statistical approaches, the performance improved when individual sites were first aggregated to country-level averages. Results suggest that statistical models, as compared to CERES-Maize, represent a useful if imperfect tool for projecting future yield responses, with their usefulness higher at broader spatial scales. It is also at these broader scales that climate projections are most available and reliable, and therefore statistical models are likely to continue to play an important role in anticipating future impacts of climate change.

1. Introduction

Improved understanding of the potential effects of climate change on crop yields is central to planning appropriate and timely responses. Analysts wishing to anticipate these effects must inevitably rely on some conceptual or numerical model of how crop yields respond to climate. A widely used approach to this prediction problem is to rely on numerical models that emulate the main processes of crop growth and development. These process-based models are typically developed and tested using experimental trials and thus offer the distinct advantage of leveraging decades of research on crop physiology and reproduction, agronomy, and soil science, among other disciplines. Yet these models also require extensive input data on cultivar, management, and soil conditions that are unavailable in many parts of the world.

More significantly, even in the presence of such data these models can be very difficult to calibrate because of a large numbers of uncertain parameters. Often this parameter uncertainty is ignored and a subjective decision is made to proceed with a single set of parameter values that produces acceptable agreement with observations. When uncertainties in parameter values are explicitly considered, however, the uncertainty estimates for model projections can widen substantially. For example, Iizumi et al. (2009) and Tao et al. (2009) describe efforts to estimate distributions of

* Corresponding author at: 473 Via Ortega, MC 4205, FSE, Stanford, CA 94305, United States.
E-mail addresses: dlobell@stanford.edu, dlobell@stanfordalumni.org (D.B. Lobell).

0168-1923/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.agrformet.2010.07.008

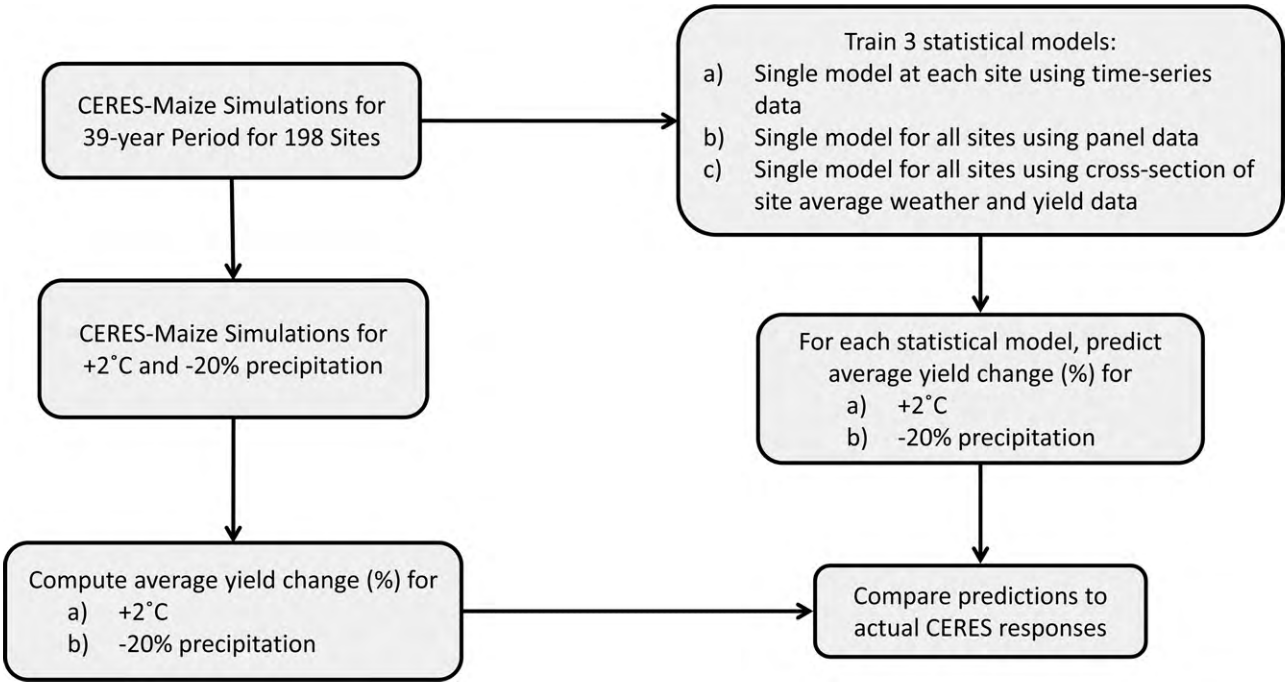

Fig. 1. A schematic outline of the perfect model approach used in this study. CERES-Maize is used to simulate "true" outcomes for variations in weather conditions over a 39-year period. Three different statistical models are then fit to these "data", and used to predict the response to scenarios of higher temperatures or reduced precipitation. Comparison with the "true" CERES response to these changes is used to measure the statistical models' abilities to predict crop responses to climate change.

parameter values for a simplified process-based model from data on yields of rice and maize, respectively. Both studies employed a Markov Chain Monte Carlo technique to retrieve parameter distributions, with the width of these distributions reflecting the inability of historical datasets to completely constrain parameter values. Parameter uncertainties then translated to large uncertainties in projecting responses to climate change, particularly for future scenarios that exceeded those in the calibration period (Iizumi et al., 2009).

Statistical models, in which historical data on crop yields and weather are used to calibrate relatively simple regression equations, provide a common alternative to process-based models. Three main types of statistical approaches are found in the literature: those based purely on time series data from a single point or area (time series methods), those based on variations both in time and space (panel methods), and those based solely on variations in space (cross-section methods). Time-series models are generally believed to have the advantage of capturing the behavior particular to the given area, whereas panel and cross-section methods must assume common parameter values for all locations, and cross-section methods in particular are prone to errors from omitted variables such as soil quality or fertilizer inputs that vary spatially. On the other hand, time-series models are often limited by data whereas panel and cross-section methods can aggregate data from multiple sites. A further discussion of the strengths and limits of particular methods in the context of predicting yield responses to climate change can be found in Lobell and Burke (2009).

The main advantages of statistical models are their limited reliance on field calibration data, and their transparent assessment of model uncertainties. For example, if a model does a poor job of representing crop yield responses to climate, this will be reflected in a low coefficient of determination (R^2) between modeled and observed quantities, as well as a large confidence interval around model coefficients and predictions. Although process-based models could in theory be accompanied with similar statistics, in practice they rarely are.

Statistical models are not without serious shortcomings, however, and in particular they are subject to problems of co-linearity between predictor variables (e.g., temperature and precipitation), assumptions of stationarity (e.g., that past relationships will hold in the future, even if management systems evolve), and low signal-to-noise ratios in yield or weather records in many locations. An example of the co-linearity problem was highlighted by Sheehy et al. (2006) in response to the statistical models of Peng et al. (2004), which showed a 10% decline of Philippine rice yields with a 1 °C increase in average minimum temperature (T_{min}). Sheehy et al. (2006) argued that solar radiation was a strong negative correlate of T_{min}, and thus an apparent negative effect of warming could easily arise from a positive effect of higher solar radiation. Similarly, Lobell and Ortiz-Monasterio (2007) showed that historical correlations between T_{min} and wheat yields in Mexico arose in part because of a negative correlation between solar radiation and T_{min}.

Despite the frequent caveats to results from statistical approaches (e.g., White, 2009), little work has been done to systematically evaluate their performance for predicting yield responses to climate. As their widespread use continues, it would be useful to know the specific conditions under which these models are most likely to mislead, and to quantify the errors incurred by adopting this convenient if imperfect approach. Moreover, because the aforementioned factors that challenge statistical approaches (e.g., co-linearity, signal-to-noise) will vary with scale, it is useful to evaluate statistical models at a range of different spatial scales.

As a step toward these goals, the current study evaluates the ability of statistical models to predict yield responses to temperature and precipitation change for nearly 200 sites in Sub-Saharan Africa. Since the "true" yield responses are unknown, we invoke the "perfect model" approach whereby a different model is used to simulate data, and a statistical model is tested for its ability to recreate the underlying relationships between climate and yields. This is a common technique, for instance, in climate modeling studies where one model is used as "observations" and the others are tested for their ability to reproduce observations (Murphy et al., 2004; Tebaldi and Knutti, 2007). This approach does not rely on

Please cite this article in press as: Lobell, D.B., Burke, M.B., On the use of statistical models to predict crop yield responses to climate change. Agric. Forest Meteorol. (2010), doi:10.1016/j.agrformet.2010.07.008

Table 1
Location of sites with unique soil profiles where CERES simulations were performed.

Country	Number of sites
Burundi	11
Botswana	22
Cameroon	12
Guinea	1
Ivory Coast	8
Kenya	27
Lesotho	14
Mali	13
Niger	11
Nigeria	1
Rwanda	6
South Africa	1
Sudan	19
Uganda	9
Zambia	34
Zimbabwe	9
Total	198

the "perfect model" actually being perfect (which no model is), but rather tests the ability of a given model and calibration technique to recreate the behavior of a reference model.

In this case, we use the well established and widely used process-based model CERES-Maize as our "perfect model" to simulate historical yields, and then fit statistical regressions to the simulated data. We then evaluate the performance of the statistical models for different sites, level of spatial aggregation, and number of years used to calibrate the model. Given limitations in CERES-Maize, the goal of this paper is not to present a final verdict on statistical models, but rather to understand more fully their general level of performance and the most influential sources of errors.

2. Methods

Fig. 1 illustrates the sequence of steps used to evaluate statistical models. The study focused on maize in Sub-Saharan Africa, a crop and region of great relevance for evaluating the impacts of climate change on global food production and food security. Maize yields were simulated using CERES-Maize (version 4.0.2.0), a commonly used process-based model for evaluating maize growth and yield responses to changes in management and environmental conditions (Jones et al., 2003). The model has been applied in numerous climate change studies, including several focused on the African continent (Jones and Thornton, 2003; Thornton et al., 2009). Although far from perfect, the model embodies a great deal of understanding of how crop yields respond to temperature and precipitation.

2.1. Simulated "observations"

We begin by selecting a wide range of sites to ensure the inclusion of different soil and climatic settings. Because CERES-Maize requires detailed information on soil properties, the study was limited to locations where soil profiles were readily available from the WISE soil database of the International Soil Reference and Information Centre (Batjes, 1995), which was reformatted for crop model applications by Gijsman et al. (2007). A total of 213 soil profiles were located within the study region, although 15 sites were in locations where simulations with CERES-Maize indicated very poor growing conditions, defined as crop failures for more than half of the years simulated. After omitting these sites from further analysis, a total of 198 sites were used in this study, with a distribution across countries as shown in Table 1.

CERES-Maize requires daily weather inputs, although measurements at this frequency over a long time period are difficult to obtain. For the purposes of this study, it was deemed sufficient to use simulated weather data. These were generated using the Marksim model (Jones and Thornton, 2000) based on the prescribed location of each site, which Marksim uses to generate daily weather. This model has been widely used and provides a realistic simulation of temperature, precipitation, and solar radiation distributions. A total of 40 years of daily weather was simulated at each site.

Other required inputs for CERES-Maize include the planting window, planting rule, maize variety, fertilizer practices, and initial soil water and nitrogen levels. The planting window at each site was based on country-specific data provided by the Food and Agriculture Organization (FAO) and other entities, as synthesized in Lobell et al. (2008). The planting rule was to sow the crop when soil moisture reached 60% of field capacity, which typically triggered planting following the second significant rain event at most African sites. This rule attempts to mimic actual farmer response to weather variations, and was deemed more realistic than simply fixing planting to a single date each year. All sites planted a generic medium-maturing maize variety.

Fertilizer was prescribed at the fairly low rate of 5 kg N ha^{-1}, which is representative of most African farms. Initial soil nitrogen levels were prescribed as 5.0 μg N/g, and initial moisture was set to 30% of the way between the soil wilting point and field capacity. Simulations at each site were independent from year to year, with the simulation for a given year beginning two months before the beginning of the site-specific planting window. The same CO_2 levels (330 ppm) were used in all simulations.

For each site, simulations were performed for the 40 years of simulated weather conditions, although since growing seasons often included January 1, only a total of 39 harvests were recorded. We refer to these simulations as "historical" or "control" as they represent current climate conditions. Two additional sets of simulations were then run for each site: one in which average temperatures were uniformly increased by +2 °C, which represents an amount of warming that is likely by mid-century, and one in which precipitation was lowered by 20%. The latter simulation corresponds to a fairly pessimistic scenario of rainfall change, as it is near the lower bound of projected precipitation changes by 2050 (Christensen et al., 2007). Neither climate change simulation is intended to represent a particular projection, but rather to evaluate yield impacts of warming and drying for changes of reasonable magnitude. The average yield impact for +2 °C was computed as the average yield for the simulations with +2 °C minus the average yields in the control, while the impact of −20% precipitation was the difference between average yields with and without −20% precipitation. Intermediate values were also tested (e.g., +1 °C) with similar qualitative results (not shown). Because average yields in the control varied widely across sites (from 0.6 to 9.1 Mg ha^{-1}) we express yield changes as % change relative to average yields in the control.

2.2. Regression analysis

With the simulations completed, we now describe the training of the statistical models as shown in Fig. 1. For each site, the 39 years of simulated data were used to fit a time-series model of the form:

$$\log(Y_t) = \beta_0 + \beta_1 T_t + \beta_2 P_t + \varepsilon_t \quad (1)$$

where Y_t, T_t, P_t are yield, growing average temperature, and growing season total precipitation, respectively, in year t, β_{0-2} represent model parameters to be fit, and ε is an error term. The values of β_{0-2} were obtained as the least-squares solution to Eq. (1). The use of growing season averages for temperature and precipitation is common in statistical approaches, although it is often criticized on the basis that aspects of sub-seasonal variations, such as long dry spells

Please cite this article in press as: Lobell, D.B., Burke, M.B., On the use of statistical models to predict crop yield responses to climate change. Agric. Forest Meteorol. (2010), doi:10.1016/j.agrformet.2010.07.008

or heat waves, can be critical to crop growth (Porter and Semenov, 2005). The ability of Eq. (1) to capture yield variability will therefore reflect, in part, the importance of climate variables other than season averages.

The 198 sites were then combined to estimate a panel regression model:

$$\log(Y_{i,t}) = \beta_{i,0} + \beta_1 T_{i,t} + \beta_2 P_{i,t} + \beta_3 T_{i,t}^2 + \beta_4 P_{i,t}^2 + \varepsilon_{i,t} \quad (2)$$

where $\beta_{i,0}$ now represents an intercept, for each site i, and squared terms for both T and P are included. These terms are typically omitted for the time-series model, as we do here, because of the limited number of observations and the fact that temperatures and precipitation span a narrow range and are therefore yield response can be reasonably approximated by a linear function (see discussion of Fig. 3 below). The site-specific intercepts are typically included in a panel regression, as we do here, to account for spatially varying factors for which no observations are present, such as soil quality. We also follow a common approach of expressing yields in log units, which assumes that a given change in T or P will have the same percent impact on yields regardless of yield level.

Finally, the average yields, T, and P are computed at each site to estimate a cross-section model:

$$\log(Y_{i,avg}) = \beta_0 + \beta_1 T_{i,avg} + \beta_2 P_{i,avg} + \beta_3 T_{i,avg}^2 + \beta_4 P_{i,avg}^2 + \varepsilon_i \quad (3)$$

where again squared terms are included for T and P to capture nonlinearities in yield response. Note that although we use similar notation for Eqs. (1)–(3), the values of the parameters β are different in each case.

Each statistical model was then used at each site to estimate the effects of a 2 °C increase in T, as well as a 20% decrease in P. The predictions from the statistical models could then be directly compared to the "true" yield responses, as measured by the yield changes simulated by CERES-Maize for these same climate changes. Thus, the ability of the statistical regression models to recreate the CERES-Maize responses of yield to climate changes represents the key measure of the fidelity of statistical models in this paper. However, we note that if actual yield responses to climate change are affected by processes not represented in CERES-Maize, such as pathogens or flooding or heat damage related to processes not in CERES, a statistical model could adequately recreate the behavior of CERES-Maize without properly capturing the true behavior of crop yields.

Care should therefore be taken when interpreting performance of statistical models in capturing CERES behavior as evidence that their yield predictions are robust. Nor should one be too quick to dismiss statistical models if they perform modestly in reproducing CERES behavior, because CERES may be omitting important factors that statistical models capture well. Finally, even if CERES-Maize perfectly captured the processes affecting crop yields, it would be difficult for an uncalibrated version of the model itself to perfectly recreate the behavior, as one would first have to determine values for the many model parameters. Nevertheless, insofar as CERES-Maize captures the basic processes underlying how maize responds to changes in temperature and precipitation, the ability of statistical models to recreate this response provides insight into the agreement between the two approaches, as well as insight into the broader usefulness of statistical approaches in estimating the impacts of climate change.

3. Results and discussion

3.1. CERES projections of yield response to climate changes

Across the 198 sites, CERES-simulated a wide range of responses to warming (+2 °C) and reduced rainfall (−20%), which reflects the diverse range of agro-environments found in Sub-Saharan Africa (Fig. 2). The median projected impact for 2 °C warming was a yield loss of 14.4%, whereas the median effect of reducing precipitation by 20% was a 5.8% yield reduction. The fact that warming of this magnitude is roughly three times more important than a 20% decline in precipitation is consistent with previous modeling studies, but perhaps counter-intuitive given the importance of rainfall in driving year-to-year changes in yield (Lobell and Burke, 2008). That is, there is widespread appreciation of the importance of precipitation for crop production because of its larger interannual variability, but temperature trends are large relative to historical variability and can (and do) become more important than precipitation trends in driving CERES-simulated yield responses to climate change. We note, however, that because each simulation year is initialized from the same conditions regardless of climate scenario, the current implementation of CERES could underestimate the effects of persistent rainfall reductions, which in reality would lower average soil moisture at the beginning of the season (Wang, 2005).

The diversity of responses to both warming and drying is consistent with previous studies that have used simulations with CERES-Maize in Africa to emphasize the spatial heterogeneity of potential impacts (Jones and Thornton, 2003; Thornton et al., 2009). As shown in Fig. 2a, much of the variation in response to temperature is linearly related to the site's average temperature in the control period, with hotter sites incurring more damage from warming. In contrast, the response to rainfall exhibits a weaker relationship with the site's average precipitation in the control period (Fig. 2c). Examination of individual simulations revealed that yield impacts for drying are closely related to the increase in water stress experience by the crop, but that this latter factor was affected by several factors in addition to average precipitation, including levels of initial nitrogen stress (not shown).

3.2. Training of statistical models on historical CERES projections

An example of the data used to train the regression models using the 40-year time series at each site is shown in Fig. 3a and b, while Fig. 3c and d and e and f illustrate the data used for the panel and cross-section models, respectively. The relationships illustrated in Fig. 3a and b were typical of many sites—a very weak and insignificant inferred effect of temperature with a positive but quite uncertain effect of rainfall. This is similar to the results of Iglesias et al. (2000), who found that CERES-simulated wheat yields at sites in Spain were closely tied to total growing season precipitation. The median R^2 for Eq. (1) across all sites was 0.17, with a range from zero to 0.69. This range of values is similar to those found in studies using actual crop and weather datasets, for instance the R^2 for the 94 crop-region combinations evaluated in Lobell et al. (2008) ranged from near zero for several crops to 0.67 for groundnuts in South Asia.

The regression model trained on the panel dataset (Eq. (2)) resulted in an R^2 of 0.54. For both temperature and precipitation, the model estimated a positive coefficient for the linear term and a negative coefficient for the squared term, consistent with the expectation of an inverted-U relationship apparent in Fig. 3c and d. The optimum average growing season temperature, where yield is maximized according to the panel regression, was 21.6 °C, with an optimum precipitation total of 830 mm. The cross-section model was fit with an R^2 of 0.59 and coefficients:

$$\log(Y_{i,avg}) = -3.5 + 0.47T_{i,avg} + 5.18 \times 10^{-4} P_{i,avg} - 0.0013T_{i,avg}^2 - 1.79 \times 10^{-7} P_{i,avg}^2$$

This inferred relationship corresponds to an optimum inferred temperature of 20.9 °C and an optimum precipitation total of 1452 mm.

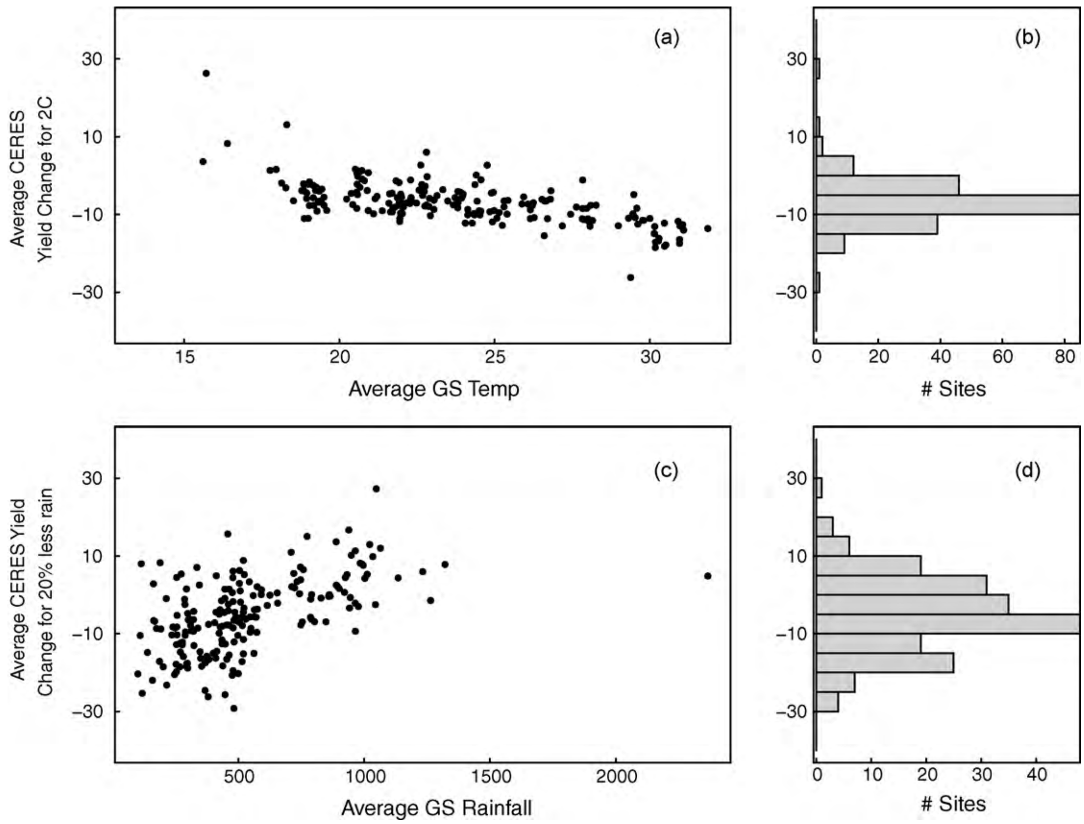

Fig. 2. Summary of yield responses to +2 °C warming and −20% precipitation as simulated by CERES-Maize. (a) Relationship between % change in average yield for +2 °C warming and average growing season (GS) temperature in the baseline period. (b) Histogram of CERES-simulated % changes in average yield for +2 °C warming. (c) Relationship between % change in average yield for −20% precipitation and average GS precipitation in the baseline period. (d) Histogram of CERES-simulated % changes in average yield for −20% precipitation.

All temperature and precipitation coefficients were significant at $p = 0.05$ for the cross-section model, and all but the precipitation squared term was significant for the panel model. As mentioned, the typical time-series model had a significant precipitation term but an insignificant temperature term. Out of 198 sites, only 22 (11%) had a temperature coefficient significant at $p = 0.05$, while 105 (53%) had a precipitation term with that level of significance. This again emphasizes the relative importance of rainfall variations for year-to-year changes in crop yields. For example, if one compares Fig. 3b and d, the range of precipitation values experienced at that one site over 40 years is nearly one-quarter the range of the entire range of precipitation simulated across all sites and years. The temperature range in Fig. 3a is, in contrast, less than 3 °C whereas the range observed across all sites and years is more than 30 °C. The relative lack of temperature variability at a single site implies that time-series models will be limited in their ability to infer temperature responses, as discussed further below.

3.3. *Projections using statistical models for individual sites*

The estimates of climate impacts from the statistical models are compared to the "true" responses of CERES in Fig. 4 and Table 2. The three statistical approaches indicated a similar median effect of +2 °C of between 11.4% and 12.7% yield loss, slightly less than but close to the "true" CERES value of 14.4%. All three approaches also agreed that the median effects of −20% precipitation would be negative but smaller than for +2 °C, although the panel model projected nearly twice the median response as CERES (−9.0% vs. −4.9%). These results relate to the bias of the statistical models, but say little about their ability to capture site-to-site differences in yield responses.

In general, the time-series models did a poor job of capturing site-to-site differences in temperature responses, but did quite well in capturing precipitation responses (Fig. 4a and d). A non-parametric measure of the scatter is the median absolute deviation (MAD), which was 14.2% for temperature responses but just 5.7% for precipitation responses (Table 2). There were no clear relationships between errors for either temperature or precipitation projections and the characteristics of sites, such as location, correlations among weather variables, or average yields. This indicates that the problems of co-linearity noted in previous work (Lobell and Ortiz-Monasterio, 2007; Sheehy et al., 2006) is not a pervasive problem, at least for the sites considered here.

Both the panel and cross-section models improved considerably on the performance of the time-series model in capturing responses to +2 °C (Fig. 4b and c). The MAD was 10.0% and 7.7% for the panel and cross-section model, respectively. As seen in Fig. 4, both the panel and cross-section models tended to give a wider distribution of predicted responses than CERES. For example, the statistical models predicted more sites with positive responses to warming than CERES, an error which occurred mainly at cooler sites. Similarly, the statistical models tended to predict more very negative impacts (>30% declines) than CERES, which occurred mainly at the warmer sites.

Unlike for temperature, the ability of statistical models to project precipitation responses did not improve for the panel and cross-section models relative to time-series models, with MAD for all three methods separated by just 1%. The panel and cross-section models did significantly change projections for many sites, but these just as often degraded rather than improved agreement with CERES values. In general, the panel and cross-section models gave very uniform impacts of precipitation across sites (Fig. 4d and e). In

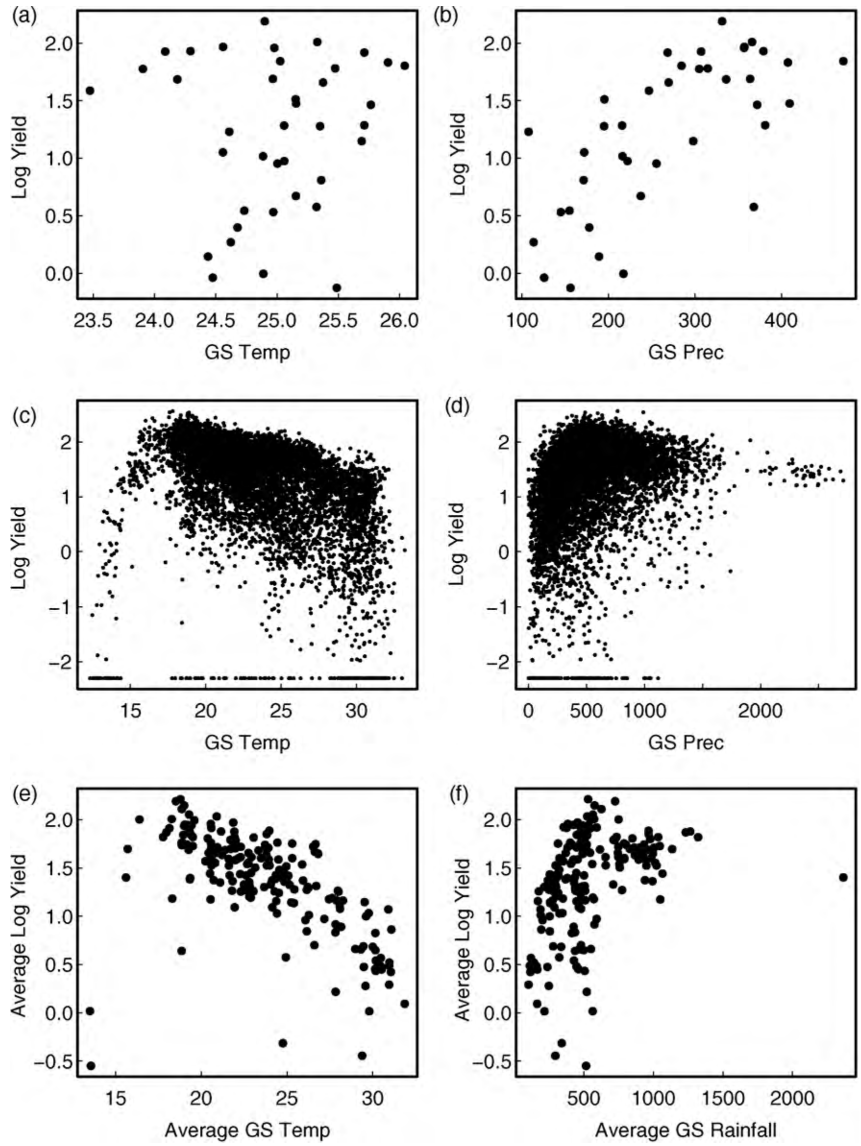

Fig. 3. (a and b) An example of simulated data used to fit the time-series model, for a site in Botswana. (c and d) The full dataset of yields vs. growing season (GS) temperature and precipitation for 39 years at 198 sites used to fit the panel model. (e and f) The full dataset of average yields vs. growing season (GS) temperature and precipitation for 198 sites used to fit the panel model.

Table 2

Summary of predicted yield changes (%) for +2 °C temperature and −20% precipitation changes from CERES models, and corresponding ability of three different statistical models trained on CERES historical simulations to predict the actual CERES response. The first four columns show values for predictions at individual field sites where CERES simulations were performed, while the right four columns show predictions where the CERES simulations were first aggregated to country averages and then used to train the statistical models. All statistics are non-parametric to avoid influence of outliers.

	Field-scale data				Country-scale data			
	CERES	Time series	Panel	Cross-section	CERES	Time series	Panel	Cross-section
Median predicted yield change (%) for +2 °C	−14.4	−11.4	−11.8	−12.7	−14.2	−13.7	−8.3	−12.3
Errors for predicting CERES response (% yield) to +2 °C								
Median error		0.5	−0.9	−2.0		−1.1	5.1	−2.2
Median absolute deviation		14.2	10.0	7.7		10.3	5.7	4.2
Rank correlation		0.33	0.67	0.67		0.13	0.81	0.81
Median predicted yield change (%) for −20% precipitation	−5.8	−3.9	−9.0	−3.4	−4.9	−2.9	−5.2	−2.6
Errors for predicting CERES response (% yield) to −20% precipitation								
Median error		1.5	−2.2	2.3		2.3	0.3	3.0
Median absolute deviation		4.8	5.6	6.7		2.9	5.4	4.1
Rank correlation		0.71	0.51	−0.48		0.76	0.37	0.42

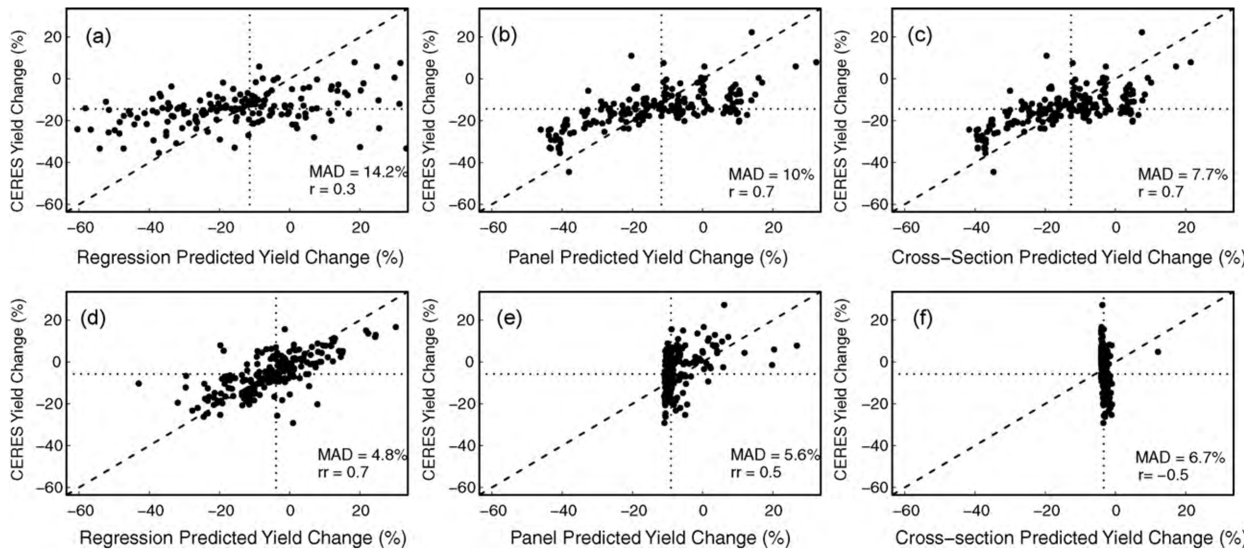

Fig. 4. A comparison of CERES-Maize predictions of average site-level yield responses (%) to 2 °C warming (*y*-axis) vs. predictions from (a) time series, (b) panel, and (c) cross-section models (*x*-axis) that were trained on historical CERES simulations. Each point represents one of 198 sites in Sub-Saharan Africa. Diagonal dashed line shows 1:1 line. Vertical and horizontal dotted lines show median yield responses for each model. (d–f) Same as (a–c) except for yield responses to 20% decline in growing season precipitation.

Table 3
The change in error measures for statistical models when trained with datasets covering fewer and fewer years. Numbers show error (% yield) in predicting CERES yield response to +2 °C warming.

Model		39 years	20 years	5 years	1 year
Time series	Median error	0.5	7.45	9.41	n/a
	Median absolute deviation	14.2	19.01	46.86	n/a
Panel	Median error	−0.9	3.38	10.61	n/a
	Median absolute deviation	10.0	7.45	10.86	n/a
Cross-section	Median error	−2.0	−2.42	−1.45	−0.89
	Median absolute deviation	7.7	8.01	8.82	10.63

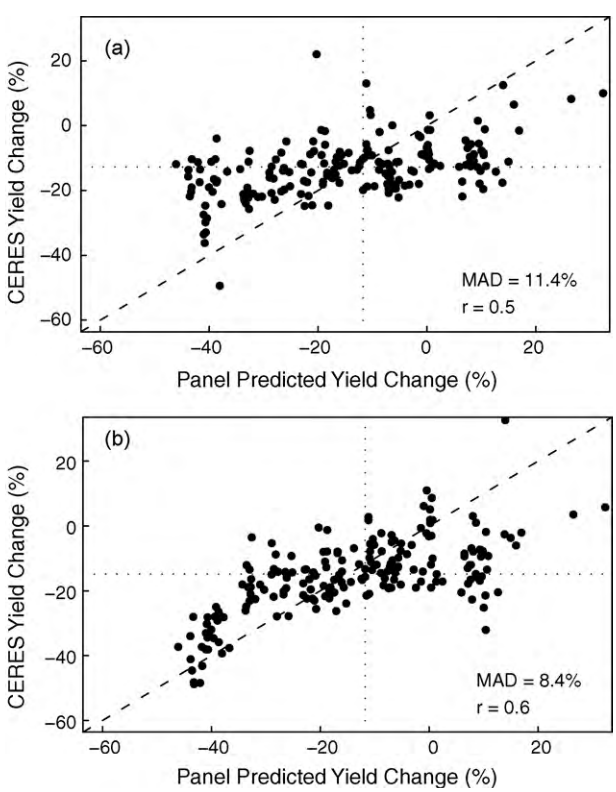

Fig. 5. Same as Fig. 4(b) except showing mean impacts for the driest (a) or wettest (b) half of years, in terms of growing season precipitation, at each site.

contrast, the CERES simulations exhibit a wide range of responses, from roughly 20% yield gains to 40% yield losses relative to baseline, for a 20% reduction in precipitation.

Thus, the results present three findings of interest. First, relative to CERES-Maize, time-series models are fairly reliable tools for projecting responses to rainfall changes, but are of limited value for anticipating temperature responses. Second, statistical models trained with spatial variation – whether panel or cross-section data – do a fairly good job at projecting temperature response but do not improve upon time-series models for projecting rainfall responses. This indicates that the added sample size and range of temperature variation that occurs when expanding to panel or cross-section models is well worth the potential penalty of restricting all sites to obey the same functional relationship, whereas the same is not true for precipitation. In part this is because temperature variations in time are much more limited than precipitation, and in part because precipitation responses are not as closely tied to average precipitation as are temperature responses to average temperature (Fig. 2).

Third, the results indicate that even for the best performing statistical models, there is considerable scatter between the projected and “true” CERES responses to climate change. This represents the fundamental inability of simple statistical models to capture the complexities of dynamic cropping systems. Even adding sub-growing season measures of weather did little to improve the ultimate performance of these statistical models. In particular, the regressions were rerun using three variables for both temperature and precipitation, corresponding to averages over each third of the growing season. The resulting regressions exhibited no significant improvements in predicting CERES responses to either temperature or precipitation changes.

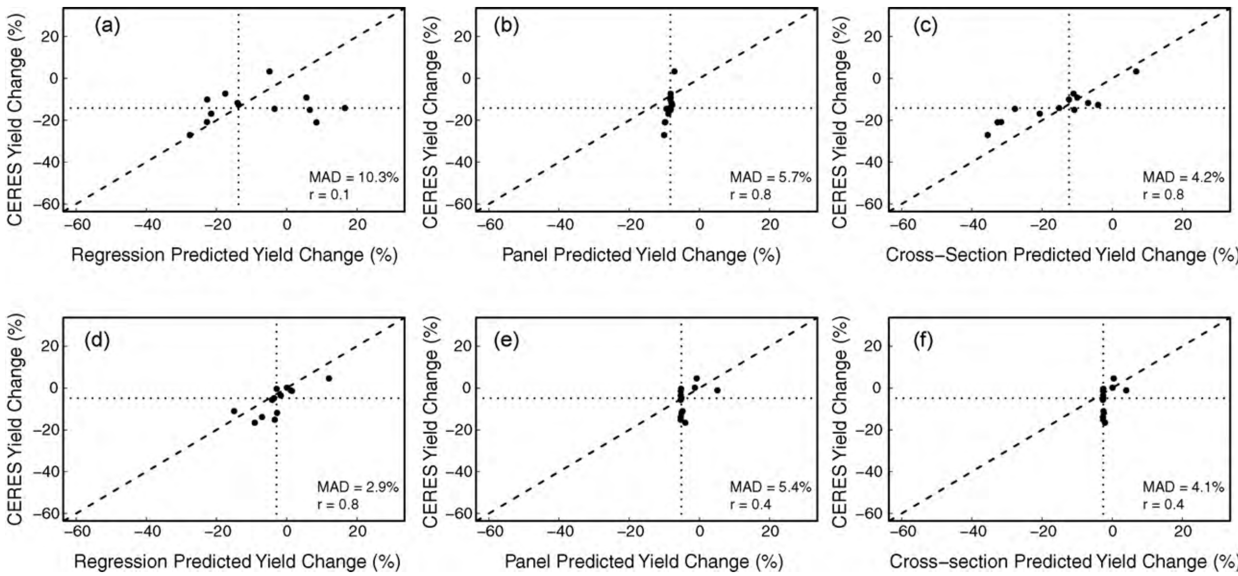

Fig. 6. A comparison of CERES-Maize predictions of average country-level yield responses (%) to 2 °C warming (*y*-axis) vs. predictions from (a) time series, (b) panel, and (c) cross-section models (*x*-axis) that were trained on historical CERES simulations aggregated to the country level. Each point represents one of 13 countries in Sub-Saharan Africa. Diagonal dashed line shows 1:1 line. Vertical and horizontal dotted lines show median yield responses for each model. (d and f) Same as (a–c) except for yield responses to 20% decline in growing season precipitation.

In addition to the effects of climate change on *average* crop yields, the impacts in particularly bad years are also of interest. Indeed, an often cited strength of process-based models is their ability to capture important interactions between variables. To examine the performance of statistical models in this respect, at each site we computed the model errors for years among the coolest vs. warmest half of the record, as well as among years with the least vs. most precipitation. The results indicated that all three statistical models tended to overestimate yield losses from warming in drier years, but there was little difference between cool and warm years (an example for panel models shown is in Fig. 5). In dry years, CERES tended to show very little effect of warming since yields were already low because of moisture stress, but statistical models did not capture this interaction with water availability. For precipitation, the performance of statistical models appeared similar across years with different temperature or precipitation amounts.

The inability of statistical models to perfectly capture CERES behavior is of course not surprising, given the far greater number of parameters and the dynamic nature of the CERES simulations, but the statistical models clearly do a decent job of approximating the "true" response. An important remaining question is whether the errors in the statistical model predictions persist when considering larger spatial scales than individual sites. That is, do the errors tend to cancel out at scales at which statistical models are more commonly applied, such as provincial or country averages? If so, the attractiveness of statistical models may depend on the scale at which projections are desired.

3.4. Projections using statistical models for sites averaged by country

To evaluate the scale-dependence of statistical model performance, the procedure outlined in Fig. 1 was repeated, except that simulations were first averaged for each year for all sites within a country. Three countries with only a single site (Table 1) were omitted, leaving 13 country aggregates for each year. The statistical models were then trained on the country aggregates, and used to predict the country-level response to warming or rainfall reductions. This is more akin to the scale at which statistical models are typically applied—for countries or regions rather than individual fields.

At the country scale, the accuracy of the statistical models for projecting impacts of warming, as measured by MAD, improved for all three statistical models (Fig. 6 and Table 2). The MAD for time-series models fell from 14.2% to 10.3%, for panel models from 10.0% to 5.7%, and for cross-section models from 7.7% to 4.2%. Results also improved in all three models for projecting impacts of precipitation reductions, although the differences in MAD were less than 2% in all cases.

The reduction of errors when aggregating to larger scales indicates that the relationship between weather and yields is more appropriately described by simple functions at coarse scales than at finer scales, an observation consistent with previous work on crop yields (Challinor et al., 2005; Hansen and Jones, 2000; Landau et al., 2000; Lobell and Field, 2007) and a broader literature on scale and environmental modeling (Addiscott and Tuck, 2001; Beven, 2002). Intuitively, this happens because many of the errors at individual fields are independent and therefore cancel out when aggregating to larger scales. While some questions related to agricultural adaptation to climate change may require projections with considerable spatial detail, particularly in topographically diverse regions (Jones and Thornton, 2003), the response of aggregated production over broad regions is of relevance to many policy questions, such as whether national food production is at risk from climate change. Moreover, while statistical models may be limited for fine-scale responses, it is often questionable whether even perfect crop models would have the necessary inputs to project impacts at these scales, since climate projections at field scales are hard to obtain and extremely uncertain (e.g., Hansen and Indeje, 2004).

Since at most 34 sites were simulated for any single country, the country aggregates do not reflect a real-world country in which thousands of individual fields contribute to total maize production. It is likely that errors from individual sites would be even further reduced when aggregating over a sample size typical of most countries. On the other hand, weather at each site was simulated independently in this study, so that in any single year nearby stations could have very different weather. In the real world, weather is highly correlated among nearby fields, so that errors would likely cancel much more slowly than if they were truly independent.

Thus, we note that care should be taken to relate the specific scale-dependence discussed above to real world simulations. To do so would require further work that characterizes the number and similarity of fields in a more realistic fashion.

3.5. Sensitivity to size of training dataset

A final factor considered in this study is the length of records available to train statistical models. This is of interest, for example, in countries where it can be extremely difficult to find reliable data prior to the 1980s. In such cases it is common to have only 20 years with which to examine relationships between yields and climatic conditions. Table 3 compares the baseline results using 39 years for 198 individual sites with shortening the period to 20 or, as an extreme, 5 years. The time-series model shows an expected increase in both median error and MAD for predicting CERES response to +2 °C. The median error jumps from near zero to over 7%, while the MAD for 20 years is only 5% greater than in the baseline case.

The panel and cross-section models show less sensitivity to number of years, as expected since they rely on spatial as well as temporal information. The cross-section model in particular is robust even with only 5 years of data to compute average yields and weather at each site. Using only a single year of data for 198 sites, the MAD rises slightly to 10.6%. Thus efforts to infer climate sensitivities using cross-sections from a survey in a single year may find reasonable results, for instance the recent work on relating farm revenue to climate in Africa (Dinar et al., 2008; Kurukulasuriya et al., 2006), although these studies focus on farm revenue rather than yield, with the former arguably more difficult to measure accurately and more sensitive to off-field decisions such as utilization of stored grains.

4. Summary and conclusions

Statistical models based on temporal or spatial variation in crop yields, or a combination of the two, are now widely used to investigate the effects of recent and future climate changes on crop yields. Although the relative strengths and weaknesses of different statistical approaches are often discussed, little has been done to systematically evaluate their ability to generate accurate projections of crop response across a range of factors that could influence their performance. This study provides a step in that direction by treating output from a process-based crop model, CERES-Maize, as observations. The advantage of using this perfect model approach is that the “true” response to climate change can be calculated, whereas with real data the future response in unknown. As discussed in the Introduction, this approach is limited in part by the fact that CERES-Maize only represents some of the many processes that affect yields.

The results provide several insights, some which were fairly surprising. All three approaches exhibit relatively little bias when trained with the full dataset (39 years at 198 sites) and used to project impacts to warming or precipitation reductions, where bias was measured non-parametrically as the median error. For example, the median error for all three methods was less than 2% for projecting impacts of +2 °C, which was much smaller than the median projected impact of −14.4%. However, the time series approach exhibited high model variance, with a lot of scatter around the 1:1 line (Fig. 3a) and a median absolute deviation above 14%. An option to reduce this variance that could be explored in future work is the use of Bayesian methods, such as ridge regressions, that shrink parameter estimates toward a prior distribution (Hastie et al., 2001; Iizumi et al., 2009). These techniques effectively add a little bias in exchange for larger reductions in variance, ultimately reducing the prediction errors.

The panel regressions, which combine all data into a single regression that includes site-specific intercepts to account for omitted time-invariant variables, proved more robust than time-series models for predicting temperature responses. This is consistent with the recent findings of Schlenker and Lobell (2010), who found that a panel regression using country yield statistics in Africa resulted in much narrower error bars than the time series approach used previously with the same dataset. Surprisingly, however, there was no visible improvement in the current study when panel models were used to project precipitation responses (Fig. 3d and e). This finding was attributed to the facts that precipitation responses in CERES showed only a weak correlation with average precipitation, and that variations in precipitation through time at a single site were large enough to fit a reliable time-series model. Cross-section models provided the most accurate predictions of temperature response, but also the poorest predictions of rainfall response. Again, this is related to the fact that precipitation response in CERES was influenced by factors such as rainfall timing and nitrogen stress that were not captured in the cross-section model.

The performance of statistical models may also be affected by several factors not considered here. One is the presence of noise in measurements of yield or weather, which could obscure the relationship between the two. Another factor is potential differences in the varieties sown in different locations. Whereas the simulations used here assumed a fixed variety, varieties often differ across space in response to local climate and other factors. Projections from cross-section models trained on such data would therefore assume that farmers are capable of automatically adapting the varieties they grow to new climate conditions, even if the new climate conditions have never experienced before at that location. Recent work that compared cross-section, panel, and time-series models for maize, soybean, and cotton production in the United States suggests that the impact of these adaptations are modest (Schlenker and Roberts, 2009). Specifically, the authors found very small differences between the three models in predicting response to high temperatures, indicating that the scope of existing adaptations to high temperatures has been small.

The performance of statistical models could also be affected by changes in climate variability, which in this study was assumed to be the same in future as in current climate. The crop, region, and type of management system considered (e.g., in this study, rainfed maize with low fertilizer inputs) also will likely influence the utility of statistical approaches. Further work is therefore needed to test the generality of conclusions across different crops and growing conditions. Perhaps most importantly, the results of the study are specific to CERES, and future work should consider other process-based models.

Overall the results emphasize three important points that users of crop models should consider in future work. First, statistical models represent a very useful if imperfect tool for projecting climate responses, with all three statistical approaches able to reproduce some of the key aspects of the simulated responses to temperature and precipitation changes. Second, the relative performance of statistical models will depend on the response in question. Time-series models appear particularly good at estimating precipitation responses, while panel or cross-section methods appear more reliable for temperature responses. Finally, the accuracy of statistical approaches depends on the spatial scale of the training data and the scale at which output projections are required. In general, statistical models appear to become more appropriate as the scale of interest becomes broader. It is also at these broader scales that climate projections are most available and reliable, and therefore statistical models are likely to continue to play an important role in anticipating future impacts of climate change. These strategies should also consider factors not addressed in the current study, such as fertilization effects of higher atmospheric CO_2 levels.

Acknowledgments

This work was supported by a grant from the Rockefeller Foundation and by NASA New Investigator Grant No. NNX08AV25G to DL. We thank Paul Switzer for helpful discussions and three anonymous reviewers for helpful comments.

References

Addiscott, T.M., Tuck, G., 2001. Non-linearity and error in modelling soil processes. European Journal of Soil Science 52 (1), 129–138.
Batjes, N.H., 1995. A Homogenized Soil Data File for Global Environmental Research: A Subset of FAO, ISRIC and NCRS Profiles (Version 1.0). Wageningen, the Netherlands.
Beven, K., 2002. Towards a coherent philosophy for modelling the environment. Proceedings of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences 458 (2026), 2465–2484.
Challinor, A.J., Wheeler, T.R., Slingo, J.M., Craufurd, P.Q., Grimes, D.I.F., 2005. Simulation of crop yields using ERA-40: limits to skill and nonstationarity in weather–yield relationships. Journal of Applied Meteorology 44 (4), 516–531.
Christensen, J.H., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, I., Jones, R., Kolli, R.K., Kwon, W.-T., Laprise, R., Rueda, V.M., Mearns, L., Menéndez, C.G., Räisänen, J., Rinke, A., Sarr, A., Whetton, P., 2007. Regional climate projections. In: Solomon, S., et al. (Eds.), Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom/New York, NY, USA.
Dinar, A., Somé, L., Hassan, R., Mendelsohn, R., Benhin, J., 2008. Climate Change and Agriculture in Africa: Impact Assessment and Adaptation Strategies. Earthscan/James & James.
Gijsman, A.J., Thornton, P.K., Hoogenboom, G., 2007. Using the WISE database to parameterize soil inputs for crop simulation models. Computers and Electronics in Agriculture 56 (2), 85–100.
Hansen, J.W., Indeje, M., 2004. Linking dynamic seasonal climate forecasts with crop simulation for maize yield prediction in semi-arid Kenya. Agricultural and Forest Meteorology 125 (1–2), 143–157.
Hansen, J.W., Jones, J.W., 2000. Scaling-up crop models for climate variability applications. Agricultural Systems 65 (1), 43–72.
Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics. Springer, New York, 533 pp.
Iglesias, A., Rosenzweig, C., Pereira, D., 2000. Agricultural impacts of climate change in Spain: developing tools for a spatial analysis. Global Environmental Change 10 (1), 69–80.
Iizumi, T., Yokozawa, M., Nishimori, M., 2009. Parameter estimation and uncertainty analysis of a large-scale crop model for paddy rice: application of a Bayesian approach. Agricultural and Forest Meteorology 149 (2), 333–348.
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., Ritchie, J.T., 2003. The DSSAT cropping system model. European Journal of Agronomy 18 (3–4), 235–265.
Jones, P.G., Thornton, P.K., 2000. MarkSim Software to generate daily weather data for Latin America and Africa. American Society of Agronomy, 445–453.
Jones, P.G., Thornton, P.K., 2003. The potential impacts of climate change on maize production in Africa and Latin America in 2055. Global Environmental Change-Human and Policy Dimensions 13 (1), 51–59.
Kurukulasuriya, P., Mendelsohn, R., Hassan, R., Benhin, J., Deressa, T., Diop, M., Eid, H.M., Fosu, K.Y., Gbetibouo, G., Jain, S., 2006. Will African Agriculture Survive Climate Change? The World Bank Economic Review 20 (3), 367.
Landau, S., Mitchell, R.A.C., Barnett, V., Colls, J.J., Craigon, J., Payne, R.W., 2000. A parsimonious, multiple-regression model of wheat yield response to environment. Agricultural and Forest Meteorology 101 (2–3), 151–166.
Lobell, D.B., Burke, M.B., 2008. Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation. Environmental Research Letters 3 (3), 034007.
Lobell, D.B., Burke, M.B. (Eds.), 2009. Climate Change and Food Security: Adapting Agriculture to a Warmer World. Springer.
Lobell, D.B., Field, C.B., 2007. Global scale climate–crop yield relationships and the impacts of recent warming. Environmental Research Letters 2, 004000 (7 pp.).
Lobell, D.B., Ortiz-Monasterio, J.I., 2007. Impacts of day versus night temperatures on spring wheat yields: a comparison of empirical and CERES model predictions in three locations. Agronomy Journal 99 (2), 469–477.
Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L., 2008. Prioritizing climate change adaptation needs for food security in 2030. Science 319 (5863), 607–610.
Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth, D.A., 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430 (7001), 768–772.
Peng, S., Huang, J., Sheehy, J., Laza, R., Visperas, R., Zhong, X., Centeno, G., Khush, G., Cassman, K., 2004. Rice yields decline with higher night temperature from global warming. Proceedings of the National Academy of Sciences of the United States of America 101 (27), 9971–9975.
Porter, J.R., Semenov, M.A., 2005. Crop responses to climatic variation. Philosophical Transactions of the Royal Society B-Biological Sciences 360 (1463), 2021–2035.
Schlenker, W., Lobell, D.B., 2010. Robust negative impacts of climate change on African agriculture. Environmental Research Letters, 014010 (8 pp.).
Schlenker, W., Roberts, M.J., 2009. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. Proceedings of the National Academy of Sciences 106 (37), 15594–15598.
Sheehy, J.E., Mitchell, P.L., Ferrer, A.B., 2006. Decline in rice grain yields with temperature: models and correlations can give different estimates. Field Crops Research 98 (2–3), 151–156.
Tao, F., Yokozawa, M., Zhang, Z., 2009. Modelling the impacts of weather and climate variability on crop productivity over a large area: a new process-based model development, optimization, and uncertainties analysis. Agricultural and Forest Meteorology 149 (5), 831–850.
Tebaldi, C., Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate projections. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 365 (1857), 2053–2075.
Thornton, P.K., Jones, P.G., Alagarswamy, G., Andresen, J., 2009. Spatial variation of crop yield response to climate change in East Africa. Global Environmental Change 19 (1), 54–65.
Wang, G., 2005. Agricultural drought in a future climate: results from 15 global climate models participating in the IPCC 4th assessment. Climate Dynamics 25 (7–8), 739–753.
White, J.W., 2009. Comments on a report of regression-based evidence for impact of recent climate change on winter wheat yields. Agriculture, Ecosystems and Environment 129 (4), 547–548.

Please cite this article in press as: Lobell, D.B., Burke, M.B., On the use of statistical models to predict crop yield responses to climate change. Agric. Forest Meteorol. (2010), doi:10.1016/j.agrformet.2010.07.008