Paper:
Watson, J and Challinor, AJ (2013) *The relative importance of rainfall, temperature and yield data for a reginal-scale crop model.* Agricultural and Forest Meteorology, 170. 47 - 57. ISSN 0168-1923

http://dx.doi.org/10.1016/j.agrformet.2012.08.001

The relative importance of rainfall, temperature and yield data for a regional-scale crop model

Authors

James Watson[1] and Andrew Challinor

Affiliation and Postal Address

Institute for Climate and Atmospheric Science
School of Earth and Environment
University of Leeds, United Kingdom
LS2 9JT

Keywords

Crop models; Climate models; Climate variability; Uncertainty; Crop yield; Data quality.

Abbreviations

a. GLAM – General Large-Area Model for Annual Crops
b. RMSE – root mean square error
c. RMSD – root mean square difference
d. YGP – yield gap parameter
e. TE – transpiration efficiency
f. *p* – bias rate
g. *r* – correlation coefficient of projected yield and observed yield

[1] Corresponding author. Ph: +44 113 343 4931; Fax: +44 113 343 5259.

Abstract

When projecting future crop production, the skill of regional scale (> 100km resolution) crop models is limited by the spatial and temporal accuracy of the calibration and weather data used. The skill of climate models in reproducing surface properties such as mean temperature and rainfall patterns is of critical importance for the simulation of crop yield. However, the impact of input data errors on the skill of regional scale crop models has not been systematically quantified. We evaluate the impact of specific data error scenarios on the skill of regional-scale hindcasts of groundnut yield in the Gujarat region of India, using observed input data with the GLAM crop model. Two methods were employed to introduce error into rainfall, temperature and crop yield inputs at seasonal and climatological timescales: (1) random temporal resequencing, and (2) biasing values.

We find that, because the study region is rainfall limited, errors in rainfall data have the most significant impact on model skill overall. More generally, we find that errors in inter-annual variability of seasonal temperature and precipitation cause the greatest crop model error. Errors in the crop yield data used for calibration increased Root Mean Square Error by up to 143%. Given that cropping systems are subject both to a changing climate and to ongoing efforts to reduce the yield gap, both potential and actual crop productivity at the regional scale need to be measured.

We identify three key endeavours that can improve the ability to assess future crop productivity at the regional-scale: (i) increasingly accurate representation of inter-annual climate variability in climate models; (ii) similar studies with other crop models to identify their relative strengths in dealing with different types of climate model error; (iii) the development of techniques to assess potential and actual yields, with associated confidence ranges, at the regional scale.

1. Introduction

All projections of the impacts of climate change on crop yield rely on models. Since such models are incomplete representations of complex biological processes, their accuracy is limited by their structure. Model accuracy is also limited by error (i.e. inaccuracy) and uncertainty (known imprecision) in model inputs. Projections of crop yield using crop and climate models have identified uncertainty in climate as a significant, if not dominant, contribution to total projected uncertainty (e.g. Baron et al. 2005; Challinor et al., 2010, 2009a, 2005a; Cruz et al. 2007; Mearns et al. 2003; Trnka et al., 2004). Lobell [this issue] finds that ignoring measurement errors when using an empirical crop model can underestimate sensitivity to rainfall by a factor of two or more. These sensitivities have clear implications for assessments of the impact of climate change on food production and food security, and for the way in which adaptation options are formulated (e.g. Challinor, 2009).

Calibration and weather inputs can have random or systematic errors at a variety of spatial and temporal scales. For example, climate models can overestimate the number of rainy days whilst underestimating rainfall intensity (Randall et al., 2007) and may also fail to represent the sub-seasonal variation in rainfall. Observational data such as crop production or daily weather may contain uncorrelated, random errors introduced in measurement or recording, and systematic bias from aggregation to the regional scale. These biases each have different implications for crop simulation, with some types of error being easier to correct than others (Challinor et al., 2005b).

While current efforts are underway to both quantify and reduce uncertainty in climate models (e.g., the Coupled Model Intercomparison Project Phase 5), the specific impact of such errors on crop models at the regional scale are still unknown. Crop models are often calibrated using historical crop yield data, which is made available at the regional scale by organizations such as the Food and Agriculture Organization of the United Nations (FAO) and the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). Unlike climate model output, the quality and availability of this data varies region by region.

The sensitivity of field-scale crop models to weather inputs has been assessed in a number of studies (e.g. van Bussel et al., 2011; Dubrovsky et al., 2000), and the importance of calibration data in field scale models has previously been analyzed (Batchelor et al., 2002). Sensitivity studies with regional scale crop models, such as those reviewed by Challinor et al. (2009b), are less common. These models integrate inputs at different scales, and are effectively test beds of theory about what processes dominate variability in crop yield at these scales. Regional-scale models tend to be less complex than field-scale models, therefore the impact of errors in input data on these two types of model can be expected to differ. Berg et al. (2010) investigated the sensitivity of a large-scale crop model to errors in rainfall inputs, but to date errors in rainfall, temperature and yield observations have not been systematically studied at this scale.

Our objective in this study is to quantify the contribution made by specific data error scenarios to error in regional scale yield projections. We take a published set of crop yield simulations (Challinor et al., 2004), and introduce error into the input data in order to assess its impact on model error. We use a set of simulations where regional-scale yields were reproduced skillfully using observed weather data. The errors introduced to that data can be understood to represent uncertainty in the simulation of weather by a climate model, and errors in the collection and collation of crop yield information. The errors are introduced (i.e. simulated) at a range of temporal scales and using two methods. The first samples and resequences values from the baseline climate to break temporal structure (described in Section 2.2.1). The second alters the observed values such that they include and then exceed observed values from the baseline climate (Section 2.2.2). Results from applying these methods to rainfall, temperature and yield inputs are presented in Section 3.1, and a comparison of these two schemes is given in Section 3.2. Three model configurations are used in the study. These are described in Section 2.1 and the difference in results between these model configurations is described in Section 3.3. The implications of the results for regional scale crop modelling are discussed in Section 4 and conclusions are drawn in Section 5.

2. Material and Methods

2.1 Crop model

Crop yields were simulated using the General Large Area Model for annual crops (GLAM). This model, which is freely available for non-commercial use via a licence agreement, has been used to simulate the mean and variability of yields in current and future climates across the tropics (see Challinor et al., 2010).

GLAM uses soil properties, a planting window, rainfall, solar radiation and minimum and maximum temperature to simulate crop growth and development on a daily time step. It is calibrated by adjusting the Yield Gap Parameter (YGP) to minimise discrepancies (as measured by Root Mean Square Error, RMSE) between simulated and observed yields. Altering YGP alters the rate of change of leaf area index with respect to time. We replicated the simulations of Challinor et al. (2004), hereafter referred to as C2004. C2004 simulated groundnut yield across India on a 2.5 by 2.5 degree grid for the period 1966 to 1989, using observed annual yield data for calibration and gridded weather data derived from observations. Rainfall data were daily; the monthly temperature data were linearly interpolated to produce the daily values required by GLAM. Solar radiation data were monthly climatological solar radiation, which were linearly interpolated to daily values. Using these data, C2004 were able to demonstrate the importance of both inter-annual and intra-seasonal variability in rainfall in determining crop yield. The parameter set of C2004 is based on literature searches to identify plausible ranges of parameters and subsequent minimisation of RMSE. All model parameters lie near the centre of the ranges, with the exception of transpiration efficiency (TE). The parameter set has subsequently been used and tested extensively (Challinor et al., 2009a, 2007, 2005a,b,c; Challinor and Wheeler 2008a,b).

The analysis presented here focuses on a grid cell in Gujarat (GJ) in which both the observed inter-annual variability in yield and the skill of the crop model in reproducing that variability was high (correlation coefficient, r=0.74). The RMSE of the GJ grid cell is higher than that of the other two grid cells examined in detail by C2004 (281 kg ha^{-1} as compared to 105 and 176 kg ha^{-1}). However, both of these grid cells, and many of the others simulated in C2004 had lower inter-annual variability in yield than GJ, and also a lower correlation coefficient between observed and simulated yield. Thus the choice was made to focus on a grid cell where GLAM has demonstrable skill in reproducing inter-annual variability, despite the absolute RMSE not being the lowest.

The replicated C2004 simulations for GJ were used as a control experiment. In all cases, unless otherwise reported, the model was calibrated by varying YGP in steps of 0.05, from a minimum of 0.05 to a maximum of 1. The calibrated value of YGP is that which produces the lowest RMSE over the whole time period. Note that variation in model skill when calibration and evaluation time periods were separated was assessed by C2004, and found to be small. Replication of the original results was not perfect, due to minor modifications made to the model code since 2004. C2004 reported a model yield RMSE in GJ of 281 kg ha^{-1}, while the control simulation in the current study gave a yield RMSE of 274 kg ha^{-1}. Model skill was measured by two metrics: RMSE, and the correlation coefficient of projected yield and observed yield (r).

Three crop model configurations were used in the study. Configuration A is taken directly from C2004 and reproduces the yields from that study (subject to the minor differences noted above). Configuration B is identical to A, but with the GLAM high temperature stress parameterisation of Challinor et al. (2005c) activated. This configuration was used because the temperatures resulting from some of the biases described in Section 2.2.2 below fall outside the range observed in the baseline climate. In particular, they exceed the critical value beyond which anthesis and pod set are affected. Configuration C is identical to B, but with the transpiration efficiency (TE) set to 2.5 Pa. This new value is at the centre of the range of values identified from the literature by C2004. Since TE is the only model parameter not found by C2004 to be near the centre of the range suggested by the literature, configuration C approximates a set of simulations where little *a priori* calibration of the model was carried out. In these simulations, the primary impact of the use of yield data is through the calibration parameter, YGP. Therefore, through comparing configurations A and C, conclusions may be drawn on the importance of historical crop yield data in the development of model parameterisations.

2.2 Simulating model input errors

Rainfall, temperature and yield model inputs were each perturbed using two methods. Random temporal resequencing (referred to concisely as shuffling) of the primary data (daily rainfall, monthly temperature and annual crop yield) was used to simulate errors where certain temporal information is destroyed, but values remain consistent with the current climate. The second method biased the primary input data across a range that includes, and also exceeds, values found in the baseline climate. These two methods are described in detail in Sections 2.2.1 and 2.2.2. Both of these methods were used to assess the impact of data errors on GLAM at three timescales: subseasonal, seasonal and climatological. Minimum and maximum temperature values were perturbed simultaneously in order to maintain consistency in the diurnal temperature range. Since we are interested in the effect of errors in the input data available to the model, only the relevant observations were perturbed – not the interpolated values.

2.2.1 Random shuffling of input data

In order to assess the importance of temporal information in GLAM's input given the current climate, values from the full June to September, 1966 to 1989, dataset were randomly shuffled at three timescales, according to the following three operations:

1. *Shuffle-Subseason*: daily rainfall and monthly temperature values shuffled within a season, which preserves inter-annual variability and climatology.
2. *Shuffle-Season*: rainfall, temperature and yield seasons shuffled as individual units (i.e., keeping within-season values intact), which retains subseasonal information.
3. *Shuffle-All*: subseasonal and inter-annual variability in rainfall and temperature are both altered by shuffling values across the entire dataset.

Yield calibration data were only included in Shuffle-Season, as only seasonal values exist. Each shuffling operation was repeated using 1000 unique random number seeds, so that their aggregate behaviour could be determined. GLAM was then run on each shuffled dataset in turn, where all inputs were the same as in C2004 except for a single shuffled input type (i.e., the effect of rainfall, temperature and yield shuffling were each tested separately). This procedure was repeated for parameter configurations A, B and C.

2.2.2 Biasing input data

Biasing temperature, rainfall and yield allows the simulated errors to go beyond the range of values of these variables that are observed in the current climate. Two options were considered for assessing the impact of input data biases: using known climate model error to alter observed weather, and systematically perturbing weather by introducing standardised noise. The first option has the advantage of clear links to the current skill of climate models and the second has the advantage of inter-comparability across the error introduction experiments. Since the second option

permits qualitative comparison of simulated error with existing climate model error, this method was chosen. As with the shuffling, each input data variable was perturbed in isolation, to assess its individual impact on crop model skill. Since the variables perturbed are in different units, we chose to base the bias rate *p* on standard deviation, to permit comparison across variables. Standard deviation is a commonly used aggregate variable characteristic which is in the same units as the variable being considered. Biased values were randomly chosen from the normal distribution defined by a reference value *v* and a standard deviation equal to *p*% of the standard deviation of the input values being perturbed. For example, when p = 0%, the perturbed value will equal *v*, and as *p* is increased, the likelihood of perturbed values being chosen further from *v* increase.
Datasets were perturbed at three timescales, using the following operations:

i. *Bias-Day*. Each daily rainfall value was perturbed independently of all other values. That is, each input value *v* was replaced with a perturbed value *v'* chosen from the normal distribution with a mean of *v* and a standard deviation *p*% of the climatic rainfall standard deviation. Note that the use of the term 'bias' here has been chosen to simplify the naming scheme – this operation does not uniformly alter multiple values simultaneously.
ii. *Bias-Season*: A single adjustment of value *d* was applied to all input values across the entire growing season in any one year. For rainfall and temperature inputs, *d* was chosen by subtracting the seasonal mean from the value *v'* selected from a normal distribution with mean equal to the seasonal mean, and standard deviation equal to *p*% of the seasonal standard deviation. Since only single yield values were available per season, biased yield values were calculated according to their climatological standard deviation.
iii. *Bias-Climate*: All input values were uniformly altered by the single value *d*, chosen by subtracting the climatological mean from a value chosen from the normal distribution with mean equal to the climatological mean, and standard deviation equal to *p*% of the climatological standard deviation. For temperature and precipitation, this climatological bias represents an error in the simulation of the mean climate, with no error in inter-annual variability. For yield data, it represents a systematic bias in the measurement of regional-scale crop yield data.

Each of these operations were performed for values of *p* ranging from 0 to 299, so that the impact of biases chosen from distributions with up to three times the input standard deviation were tested. As in Section 2.2.1, each perturbed variable was tested in isolation, with all other inputs the same as in C2004. GLAM was run on each biased dataset with 100 random number seeds. Figure 1 provides an example illustration of the effect of climatological biases on these GLAM runs, while Table 1 summarizes the shuffling and bias experiments performed in this study.

Table 1 Experiments performed for each input type and dataset operation. Shaded cells indicate studies that were not performed. Operations that resulted in an average RMSE that differed from the result of the baseline simulation by more than 50% are marked with a ♦.

	Shuffle Subseason	*Shuffle Season*	*Shuffle All*	*Bias Day*	*Bias Season*	*Bias Climate*
$Rain_A$		♦	♦		♦	♦
$Rain_B$		♦	♦		♦	♦
$Rain_C$						
$Temp_A$					♦	

$Temp_B$					♦	
$Temp_C$					♦	
$Yield_A$						♦
$Yield_B$						♦
$Yield_C$						♦

3. Results

An overview of the input data operations that, on average, resulted in more than 50% difference in RMSE is shown in Table 1. While this average effect on RMSE is a crude indicator of the impact of each data operator on GLAM's performance, characteristics such as GLAM's resilience to varying degrees of perturbation types, and the relative spread of behaviours across random seeds, are key to understanding the true impact of these errors at the regional scale. This section describes these results, and compares the relative impact of data operations across input variables.

3.1 Impact of temperature, rainfall and yield error on model skill

Figure 2 shows the results of shuffling, in turn, temperature, precipitation and yield in model configuration A. In the vast majority of cases, introducing error into these variables increases the RMSE of the simulated yield. The largest impact on RMSE comes from shuffling rainfall seasons. Shuffling of temperature seasons also results in RMSE that, in the vast majority of cases, is greater than that of the control simulation. Shuffling of temperature and rainfall on subseasonal timescales both result in similar changes to RMSE.

The correlation between simulated and observed yield is also plotted in Figure 2. Altering seasonal total rainfall has by far the greatest effect on *r*, with yield and temperature perturbations having the smallest effect. For both rainfall and temperature, increases in RMSE are associated with decreases in correlation. Thus the increase in RMSE is due primarily to increased error in simulating the inter-annual variability of yield, as opposed to being associated with increased error in the simulation of mean yield. In the case of perturbed yield input, there is far less evidence of any inverse relationship between RMSE and correlation coefficient. This is because the calibration parameter, YGP, affects mean yield more than it affects inter-annual variability.

The box and whiskers diagrams in Figure 2 have a smaller number of component time series for yield than for either temperature or precipitation: there are 17 unique time series of yield, compared to 1000 for Shuffle-Season of both temperature and precipitation. This is a direct result of the calibration procedure, whereby YGP is incremented in steps of 0.05; a smaller increment would result in more time series. The difference in sample size between yield and the other two variables does not affect the character of the results (see Section 3.3).

The impact of input data bias on model skill is presented in Figures 3 (RMSE) and 4 (correlation coefficient). Each figure shows the impact averaged over all 100 random seeds. For $p < 100$, rainfall biases had a greater effect on model skill, as measured by both of these metrics, than either temperature or yield biases. At these low values of *p*, daily, seasonal and climatological rainfall biases all resulted in similar RMSE. Thus, calibration provides greater compensation for errors in yield and temperature than it does for rainfall. For $p < 50$, this compensation is almost complete: temperature and yield errors have no significant impact on model skill.

For $p > 100$, seasonal biases to temperature begin to significantly affect model skill, as measured by both correlation coefficient and RMSE. This loss of skill is caused by greater inter-annual variability in crop duration, which results in inter-annual variability in yield no longer being dominated by precipitation. In contrast, climatological biases to temperature do not on average significantly affect model skill, because the calibration procedure compensates for the mean bias in temperature. Similar behaviour is seen for rainfall: climatological biases to rainfall are more easily compensated for by calibration than seasonal biases. This is particularly evident in the correlation coefficient (Figure 4); though it can also be seen in RMSE (Figure 3). The behaviour of yield biases for $p > 100$ contrasted with that of temperature: seasonal biases to yield do not affect model skill and climatological biases do. This is a direct result of the calibration procedure, which is based on yields averaged over the whole time period.

3.2 Comparison of bias and shuffle schemes

The two schemes used to introduce error in this study are not directly comparable. In order to provide some indication of the relationship between the schemes, an analysis was conducted. Climatological mean monthly temperature was computed for each perturbation scheme in turn, and for the observed data. The percentage of random number seeds that produced at least one value outside the observed range (O_R) was calculated. This was repeated for cumulative monthly precipitation. The results varied by month, variable and scheme. Shuffle-Season, by definition, produced no values outside of current climatology. Shuffle-Subseason produced relatively high values for temperature (52 to 75%, across the four months) and September rainfall (93%) and relatively low values for June, July and August precipitation (24.2, 0 and 1.3%, respectively). A similar pattern was seen for Shuffle-All.

The results for the bias scheme are too numerous to report. In general, O_R increased with increasing p. A comparison between the shuffle and bias schemes was made by incrementing p from zero upwards and noting the first value at which O_R (biased) > O_R (shuffled). For precipitation, this occurred mostly at low values of p: 0 or 1 for June to August; 35 for September Bias-Season; and the condition was not met for September Bias-Climate. For temperature, p was higher: 9-19 for Bias-Season and 180 for Bias-Climate. Whilst in many cases O_R (biased) becomes comparable to O_R (shuffled) at relatively low values of p, the variation in the values of O_R across months and variables suggests that it is impossible to determine even a guideline range of values of p which may be equivalent to the shuffled data.

A clearer distinction between the shuffle and bias schemes can be found by assessing the results qualitatively. For example, for $p < 150$, any type of rainfall bias has a greater impact on RMSE than either temperature or yield. This is consistent with the shuffle simulations at all timescales except one: Shuffle-Subseason produces a significantly lower reduction in model skill than Bias-Day. By altering seasonal totals, Bias-Day degrades model performance in a manner not seen in the equivalent shuffle simulations. The fact that shuffle simulations either maintain or destroy the temporal structure of variables is perhaps the clearest difference between this scheme and the bias scheme. The latter, at least for moderately high values of p, always destroys temporal structure.

3.3 Comparison of model configurations

Model configurations A and B, which differ only in the activation of the high temperature stress module, produced equivalent model behaviours for both the shuffled and biased operations. The temperature data used in this study are monthly, with no attempt to reproduce observed daily extremes. Thus this equivalence is not surprising. Model configurations A and C produce different results in both the shuffled and biased cases. With no biasing or shuffling of input data, the RMSE of these two configurations is 274 and 322 kg ha^{-1}, respectively. Thus RMSE increases by 17.5% when a value of transpiration efficiency from the centre of the observed range is used instead of the calibrated value. The increase in RMSE would be larger if the value of YGP were not calibrated using yield data. At *p=0* the yield gap parameter was 0.8 for the control simulation (i.e. configuration A with no bias or shuffling), and 0.2 for the corresponding simulation of configuration C. This difference is the result of calibration compensating for the higher value of TE.

The performance of the shuffled configuration C simulations is shown in Figure 5. The broad response of rainfall and temperature across the timescales is similar to that of configuration A. However, unlike configuration A, RMSE was reduced and correlation coefficient increased by subseasonal shuffling. Also, shuffling temperature both subseasonally and seasonally (i.e. Shuffle-All) produces a lower RMSE than Shuffle-Season alone. Subseasonal shuffling, on average, makes the seasonal distribution of values more uniform than observations, and therefore less realistic. These results are therefore further manifestations of incorrect model calibration.

Since yield is a calibration input, the Shuffle-Season perturbation produced a limited number of unique model results. Configuration A produced 17 unique yield projections, with RMSE of 317, 346 and 387 together accounting for 72% of the 1000 different seeds. Configuration C resulted in 2 unique yield projections – one with a RMSE of 370 (863 occurrences) and the other with RMSE of 323 (137 occurrences). This is a direct result of the calibration procedure, whereby the yield gap calibration parameter is incremented in steps of 0.05. YGP decreases between configurations A and C, in order to compensate for the higher value of transpiration efficiency. A step of 0.05 at lower values of YGP will result in greater changes in simulated yield than the same step at higher values of YGP, thus producing less unique yield time series with the higher transpiration efficiency of configuration C. In order to test whether or not the difference between the baseline RMSE of configurations A and C is an artefact of the chosen YGP increment of 0.05, these simulations were repeated with a YGP increment of 0.01 (ie 99 simulations with YGP varied between 0.01 and 1). Similar results were found: RMSE of 274 for configuration A and 318 for C, as compared to 274 and 322 respectively for a step of 0.05.

Figure 6 presents the results from the bias simulations for configuration C. As was the case for shuffle operations, the character of the response of RMSE to rainfall, temperature and yield bias errors was similar for configurations A and C. The seasonal and climatic yield biases resulted in significantly higher RMSE in configuration C compared to A at all values of *p*. For temperature and precipitation, this difference was less marked. For precipitation, the rate of increase in RMSE in response to increased *p* was higher in configuration A (Figure 3) than in C, particularly for *p* < 50.

4. Discussion

4.1 The importance of calibration data

The interaction between model configuration and errors in rainfall, temperature and yield calibration data (Section 3.3) demonstrates the importance of both crop yield data and observed weather data. Without both of these data sources, it would have been impossible to determine where the optimal value of transpiration efficiency lay. Errors resulting from this omission would then be compounded by errors in observed yield, which is also used in the calibration procedure.

The yield calibration data in this study contributed to the skill of the model in two ways: (1) selection of crop model parameters at a country scale (configuration A vs configuration C), and (2) as the basis of regional calibration. Configuration C provides an estimate of the impact on RMSE of having insufficient data to determine a value of transpiration efficiency that is appropriate for a regional-scale groundnut model in India. The increase in RMSE of 17.5% when switching to the non-calibrated value of TE demonstrates the importance of regional-scale yield data in the development of parameterisations within regional-scale crop models. This is in addition to the important role of yield data in regional calibration and evaluation of models. In the current study, the largest increase to RMSE that was induced by introducing errors to the crop yield calibration data was 143% (Bias-Climate, p=113). For comparison, the largest increase to RMSE induced by the shuffle scheme was 60%. The role of yield data for calibration is made more important by climate change, which will affect both observed yields and transpiration efficiency, as well as other regional-scale crop parameters that have not been assessed here.

If differences in RMSE across model configurations are comparable to the uncertainty in the measurement of yield, then it is impossible to conclude which configuration is the most skilful. Since the yield data do not have error bars, this comparison is difficult to make. Some indication of uncertainty in yield measurement may come from comparing datasets. The Root Mean Square Difference (RMSD) between the all-India groundnut yield data of the Food and Agriculture Organization and that of the ICRISAT data (both used in C2004) is 33 kg ha^{-1}, 4% of the mean yield of either time series. The RMSD between configurations A and C is 96 kg ha^{-1}, which is 15% of the mean yield. Comparison of these two results suggests that the difference between configurations A and C is significant. However, disagreement across datasets of observed yields is often greater than 4%. Nicklin (in preparation) has shown that the RMSD between available groundnut yield datasets in Mali vary by region and are between 83 kg ha^{-1} and 342 kg ha^{-1}.

The importance of yield data for model calibration and evaluation will likely increase as climate continues to change and as efforts to increase yields continue. These independent, but connected, drivers of crop productivity continually alter the baseline situation that crop-climate models seek to reproduce. The role of closing yield gaps in promoting food security has been noted by many authors (e.g. Lobell et al., 2009). Bhatia et al. (2006) estimate that the yield gap for groundnut varies significantly across Gujarat: 1180 to 2010 kg ha^{-1}, which is 103-175% of the mean yield across the region. Without monitoring of the yield gap, the contribution of climate variability and change to crop productivity will be impossible to determine. Without assessments of the accuracy of yield data, it is impossible to determine how much error is introduced to regional-scale crop models through the calibration procedure.

4.2 Relative importance of rainfall, temperature and yield data

The importance of weather data to crop modelling is well established. Depending on the crop and region under consideration, the relative impact of data quality of these input variables varies. Lobell and Burke (2008) found that uncertainties in temperature generally had more of an effect than uncertainties in precipitation across 94 crop-region combinations. Mearns et al. (1996) found that simulated wheat yields were sensitive to changes in both temperature and precipitation, which depended on soil characteristics. Nonhebel (1994a) found that temperature and solar radiation data errors generated up to 35% overestimation of yield. In water-limited conditions, the model was sensitive to inaccuracies in precipitation and solar radiation data, but when there was sufficient water, it was sensitive to errors in temperature and solar radiation data (1994b). Heinemann et al. (2002) found variations in simulated yield for soybean (up to 24%), groundnut (up to 13.5%), maize (up to 7.6%) and wheat (up to 2.7%) resulting from errors in rainfall observations. Berg et al. (2010) found that the frequency and intensity of rainfall, as well as cumulative annual rainfall variability, are key data features for crop models to have skill in water-limited regions. In the current study rainfall is found to be more important than temperature in simulating crop yield (Section 3.1). This is consistent with the rainfed monsoon environment in Gujarat.

A more detailed analysis of the relative importance of rainfall, temperature and yield data in this study requires some understanding of how the shuffle and bias schemes can be compared. Whilst interpretation of the shuffle experiments in bias space is not trivial (Section 3.2), some comparisons can be made. Figure 8 shows the performance of configuration A for both shuffle and bias operations at the seasonal timescale. The bias results are those with the closest mean RMSE to the corresponding shuffle simulation. Following Taylor (2001), Figure 8 illustrates the relationship between the correlation coefficient, standard deviation and RMSE of observed and simulated yields. Errors in precipitation, whether induced through random temporal resequencing (i.e. shuffling) or through biasing, produced the largest systematic difference from observed yield.

Two other differences are clear from Figure 8: for all variables (i.e. temperature, yield and rainfall) shuffling results in a lower standard deviation in yield than biasing (points 3 vs points 4 in the figure); and the use of non-calibrated TE (point 1 vs point 2 on the figure) significantly alters simulated yields. The second of these results is discussed in Section 4.1. The first result indicates an important difference between the two methods of error introduction. In all simulations, the standard deviation in yield is lower than observations; but this is particularly true of the shuffled simulations. Associated with this lower standard deviation is a lower correlation between observed and simulated yields. Thus, by directly altering the temporal structure of the rainfall, temperature or yield data, the seasonal shuffle operation has a greater impact on the skill of the model in simulating inter-annual yield variability when compared to bias operations that result in a similar RMSE.

In order to assess the implications of the results presented above for operational crop forecasting, it is necessary to compare the errors simulated here to those found in climate models. Section 4.1 briefly discusses such an analysis for yield data. In order to assess temperature and precipitation, the HadCM3 historical climate simulation of Collins et al. (2010) was analysed. Figure 9 compares the observations used in this study to the HadCM3 simulation. The seasonal cycle of monthly precipitation is captured by the climate model, but there is a significant dry bias. This is consistent with the findings of Ines and Hansen (2006). The HadCM3 temperature data are closer to observations.

It is not possible to associate a single value of p with the HadCM3 simulation. However, using observations as a reference point, some values of p that are associated with the HadCM3 run can be calculated. This was carried out as follows. Climatological mean monthly temperature was computed for the observed data, for HadCM3, and the synthetic biased data. The RMSD of the observed monthly values and those of each of the synthetic time series was calculated. The value of p that produced the RMSD closest to the RMSD of HadCM3 and observations (p_3) was recorded. The procedure was repeated for monthly cumulative rainfall. For climatological means, the resulting values of p_3 for temperature were 271 for Bias-Season and 238 for Bias-Climate. For rainfall, p_3 was 77 and 66, respectively. The low values of p_3 for precipitation are the result of the high standard deviation in the observed values (see Figure 9) that are used to scale p. When inter-annual variability in rainfall was assessed in the error metric, by repeating the entire procedure using monthly standard deviation in lieu of mean values, p_3 values were 168 and 293 for Bias-Season and Bias-Climate respectively. Taken together, these results suggest that the range of values of p used in this study is consistent with the errors observed in climate models.

4.3 Generality of results

A number of factors that are specific to the current study affect the extent of applicability of the results found. These fall into three categories: the crop model chosen, the location chosen, and the perturbation operators used. GLAM does not account for non-climatic drivers of yield. Where biotic stresses dominate, these results are likely not relevant. Also, since Gujarat is a water-limited environment, the numerical analyses presented here are only relevant for rainfed environments, where water availability is the main determinant of yield. Furthermore, the experiments of this study were designed to allow comparison of the perturbations across different input variables, but in some cases the perturbations differed across variables. For example, the distribution of values in the rainfall dataset differ from the temperature values, so the equivalent *Bias* operations can have differing effects. Ideally, we would have the same perturbation scheme applied to all variables (comparable methods) which would have the same effect wherever applied (comparable effects). With current methods, we can only choose one of these. For this study we have chosen comparable methods, since if we had employed different methods, the differences resulting from perturbations would have been due to methodological as well as numeric-specific issues like the example described above.

Despite these limitations to the generality of results, some broader conclusions are possible. In particular, the relationship between climate model bias and crop model calibration is worthy of some discussion.

Yield data are required in order to calibrate any crop model. In the current study, YGP was used as a process-based and time-independent calibration parameter to minimise RMSE between observed and simulated yields. This process can correct a significant amount of climatological bias in temperature, but is less effective for the systematic errors in yield or precipitation data in this study (Figure 3). However, for precipitation, all three Bias perturbations in this study produce more wet than dry biases; and the ability of YGP to compensate for systematic dry bias has been shown to be greater than that for wet bias (Challinor et al., 2005d). Note also that the analysis presented in this paper likely underestimates the importance of temperature, since the simulations are based on monthly interpolated data and have no representation of daily extremes. More realistic time series of daily minimum and maximum temperature may have resulted in heat stress, which would have had an influence on the RMSE of the configuration B and C simulations.

Whilst every crop model has its own equations, parameters and calibration procedure, common characteristics may be expected across models. Any aspect of climate or weather that has been proved to be an important determinant of crop yield will be an important quantity for a climate model to simulate, regardless of the crop model used. Thus the importance of seasonal rainfall for crop simulation is not specific to GLAM. Similarly, yield data are a crucial part of the calibration and evaluation of any crop model. However, differences in model formulation mean that the relative importance of temperature, precipitation and calibration data will vary between models. Many models are more complex than GLAM and therefore have a higher number of crop-specific parameters that can interact with each other. A complete treatment of these interactions is beyond the scope of this study. Here, we investigated only two parameters (YGP and TE) at the regional scale, and have therefore most likely produced a minimum estimate of the importance of interactions between calibration parameters in other crop models.

5. Conclusions: improving the skill of crop-climate simulations

The results from this study suggest that errors in the inter-annual variability of seasonal temperature and precipitation are likely to cause greater crop model error at the regional scale than systematic bias in the simulation of climate. This study is based on one crop model alone. Similar studies with other crop models would not only assess the robustness of the results, but may also identify the relative strengths of crop models in dealing with different types of climate model error.

Regional-scale yield data for crop model calibration are central to the future of crop productivity assessments. We found increases in crop model RMSE of up to 143% when the observed yield data used for calibration were perturbed. Without assessments of the accuracy of yield data, it is impossible to determine how much error is introduced to regional-scale crop models through the calibration procedure. Where possible, confidence ranges should therefore be provided with observed yield data. Ongoing efforts to close the yield gap, coupled with changes in climate and other environmental drivers, mean that the monitoring of potential yields is also crucial. Without estimates of the yield gap, the contribution of climate variability and change to crop productivity will be impossible to determine. The spatial heterogeneity in the yields of many cropping systems is significant. Thus improved measurement of actual and potential yields at the regional scale involves not only improved monitoring, but also carefully developed geo-spatial techniques.

The results of this study suggest three key endeavours for improved assessment of future crop productivity at the regional-scale: (i) increasingly accurate representation of inter-annual climate variability in climate models; (ii) similar studies with other crop models to identify their relative strengths in dealing with different types of climate model error; (iii) the development of techniques to assess potential and actual yields, with associated confidence ranges, at the regional scale.

Acknowledgments

Thanks to Ed Hawkins (University of Reading), Chris Ferro (University of Exeter), Tom Osborne (University of Reading) and Kathryn Nicklin (University of Leeds) for insightful discussion. This work was supported by the Natural Environment Research Council (NERC).

References

Baron, C., B. Sultan, M. Balme, B. Sarr, S. Teaore, T. Lebel, S. Janicot, and M. Dingkuh, 2005: From GCM grid cell to agricultural plot: scale issues affecting modeling of climate impacts. Phil. Trans. Roy. Soc. B, 1463 (360), 2095-2108.

Batchelor, W.D., Basso, B., and Paz, J.O. (2002). Examples of strategies to analyze spatial and temporal yield variability using crop models. Eur. J. of Agron., Vol. 18, Issues 1-2, pp 141-158.

Berg, A., Sultan, B. and de Noblet-Ducoudré, N. (2010). What are the dominant features of rainfall leading to realistic large-scale crop yield simulations in West Africa? Geophys. Res. Lett., 37, L05405.

Bhatia V.S., Singh Piara, Wani S.P., Kesava Rao A.V.R. and Srinivas K. (2006). Yield gap analysis of soybean, groundnut, pigeonpea and chickpea in India using simulation modeling. Global Theme on Agroecosystems Report No. 31. Patancheru 502 324, Andhra Pradesh, India: International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). 156 pp.

Challinor, A. J., T. R. Wheeler, J. M. Slingo, P. Q. Craufurd and D. I. F. Grimes (2004). Design and optimisation of a large-area process-based model for annual crops. Agr. For. Met., 124, (1-2) 99-120.

Challinor, A. J., T. R. Wheeler, J. M. Slingo and D. Hemming (2005a). Quantification of physical and biological uncertainty in the simulation of the yield of a tropical crop using present day and doubled CO2 climates. Phil. Trans. Roy. Soc. B. 360 (1463) 1981-2194

Challinor, A. J., T. R. Wheeler, J. M. Slingo, P. Q. Craufurd and D. I. F. Grimes, (2005b). Simulation of crop yields using the ERA40 re-analysis: limits to skill and non-stationarity in weather-yield relationships. J. App. Met. 44 (4) 516-531.

Challinor, A. J., T. R. Wheeler, P. Q. Craufurd, and J. M. Slingo (2005c). Simulation of the impact of high temperature stress on annual crop yields. Agr. For. Met, 135 (1-4) 180-189

Challinor, A.J., Slingo, J.M., Wheeler, T.R. and Doblas-Reyes, F.J. (2005d). Probabilistic simulations of crop yield over western India using the DEMETER seasonal hindcast ensembles. TELLUS A, 57, 498-512.

Challinor, A. J., T. R. Wheeler, P. Q. Craufurd, C. A. T. Ferro and D. B. Stephenson (2007). Adaptation of crops to climate change through genotypic responses to mean and extreme temperatures. Agr. Eco. Env., 119 (1-2) 190-204

Challinor, A. J. and T. R. Wheeler (2008a). Use of a crop model ensemble to quantify CO2 stimulation of water-stressed and well-watered crops. Agr. For. Met, 148 1062-1077

Challinor, A. J. and T. R. Wheeler (2008b). Crop yield reduction in the tropics under climate change: processes and uncertainties. Agr. For. Met., 148 343-356

Challinor, A. J. (2009). Developing adaptation options using climate and crop yield forecasting at seasonal to multi-decadal timescales. Env. Sci. Pol. 12 (4), 453-465

Challinor, A. J., T. R. Wheeler, D. Hemming and H. D. Upadhyaya (2009a). Ensemble yield simulations: crop and climate uncertainties, sensitivity to temperature and genotypic adaptation to climate change. Clim. Res., 38 117-127

Challinor, A. J., F. Ewert, S. Arnold, E. Simelton and E. Fraser (2009b). Crops and climate change: progress, trends, and challenges in simulating impacts and informing adaptation. J. Exp. Bot. 60 (10), 2775-2789. doi: 10.1093/jxb/erp062

Challinor, A. J., E. S. Simelton, E. D. G. Fraser, D. Hemming and M. Collins (2010). Increased crop failure due to climate change: assessing adaptation options using models and socio-economic data for wheat in China. Env. Res. Lett. 5 (2010) 034012

Collins M, Booth B.B.B., Bhaskaran B., Harris G., Murphy J. M., Sexton D.M.H. and Webb M.J. (2010). Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. Clim. Dyn.

Cruz, R.V., H. Harasawa, M. Lal, S. Wu, Y. Anokhin, B. Punsalmaa, Y. Honda, M. Jafari, C. Li and N. Huu Ninh (2007): Asia. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 469-506.

Dubrovsky, M., Zalud, Z. and Stastna, M. (2000). Sensitivity of ceres-maize yields to statistical structure of daily weather series. Clim. Chan., Vol. 46, Issue 4, 447-472.

Heinemann, A. B., Hoogenboom, G. and Chojnicki, B. (2002). The impact of potential errors in rainfall observation on the simulation of crop growth, development and yield. Eco. Model., Vol. 157, No. 1, 1-21.

Ines, A.V.M. and Hansen, J.W. (2006). Bias correction of daily GCM rainfall for crop simulation studies. Agr. For. Met., Vol. 138, No. 1-4, 44-53.

Lobell, D.B. and Burke, M.B. (2008). Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation. Env. Res. Lett., Vol. 3, No. 3.

Lobell, D.B., Cassman, K. G. and Field, C.B. (2009). Crop yield gaps: their importance, magnitudes, and causes. Ann. Rev. Env. Res., Vol. 34, 179-204.

Mearns, L. O., Rosenzweig, C. and Goldberg, R. (1996). The effect of changes in daily and interannual climatic variability on CERES-Wheat: A sensitivity study. Clim. Chan. Vol. 32, No. 3, 257-292.

Mearns, L. O. (Ed.) (2003). Issues in the impacts of climate variability and change on Agriculture. Applications to the southeastern United States. Kluwer Academic Publishers.

Nonhebel, S. (1994a). Inaccuracies in weather data and their effects on crop growth simulation results. I: Potential production. Clim. Res., Vol. 4, 47-60.

Nonhebel, S. (1994b). Inaccuracies in weather data and their effects on crop growth simulation results. II: Water-limited production. Clim. Res., Vol. 4, 61-74.

Randall, D.A., Wood R.A., Bony S., Colman R., Fichefet T., Fyfe J., Kattsov V., Pitman A., Shukla J., Srinivasan J., Stouffer R.J., Sumi A. and Taylor K.E. (2007). Climate Models and Their Evaluation. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M. and Miller, H.L. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Taylor, K.E. (2001). Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res., Vol. 106 (D7), 7183–7192.

Trnka, M., M. Dubrovsky, D. Semeradova, and Z. Zalud (2004). Projections of uncertainties in climate change scenarios into expected winter wheat yields. Theor. Appl. Climatol. 77, 229-249.

van Bussel, L.G.J., Müller, C., van Keulen, H., Ewert, F. and Leffelaar, P.A. (2011). The effect of temporal aggregation of weather input data on crop growth models' results. Agr. and For. Met., Vol. 151, No. 5, 607-619.

Figures

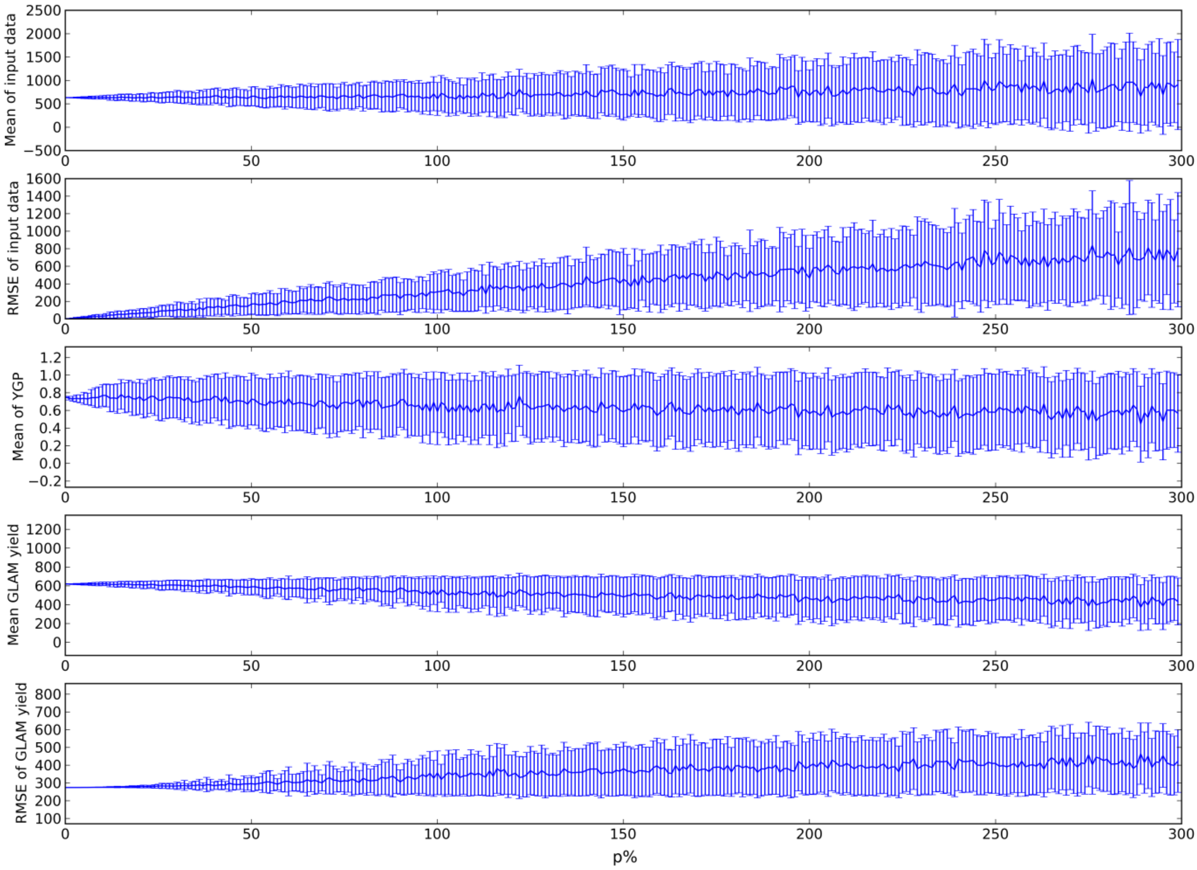

Figure 1 Effect of the Bias-Climate operation on yield inputs for GLAM configuration A. On each panel, the line represents the mean value across the 100 random number seeds used, and the bars show the standard deviation. As the value of *p* is increased, the input data values (top panel), along with the RMSE of the perturbed input data to the original observations (second panel), can be seen to deviate from the source input. The third panel plots the value of GLAM's Yield Gap Parameter (YGP), while the bottom two panels show the mean projected yield, and the RMSE of projected yield against observed yield.

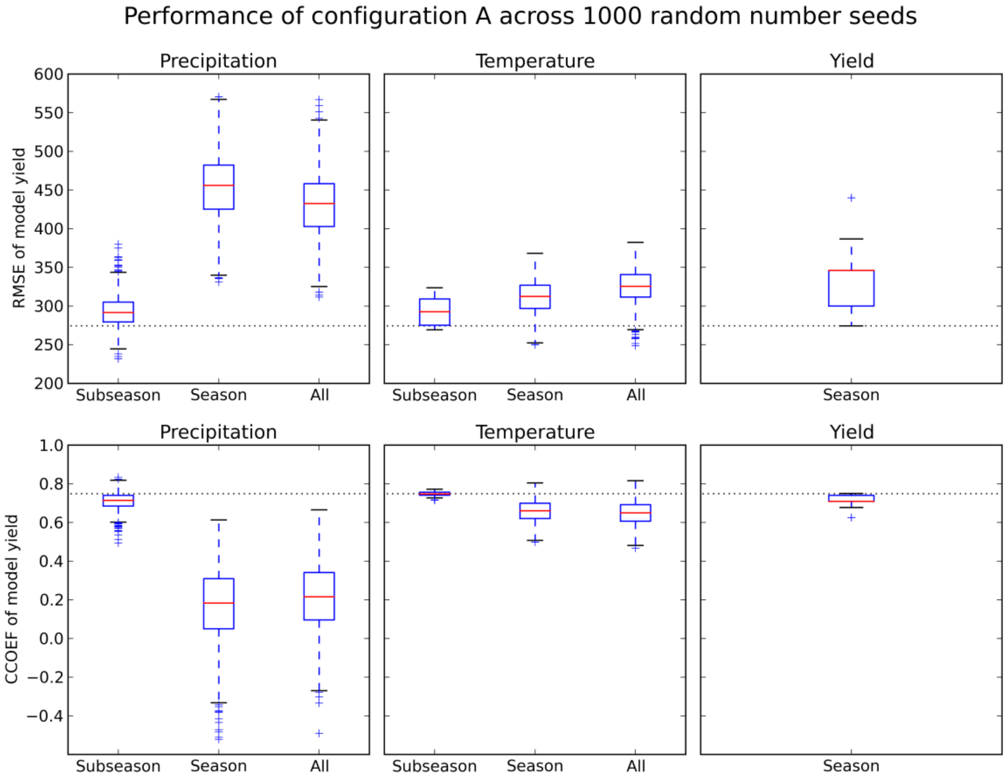

Figure 2 RMSE and correlation coefficient of observations against GLAM configuration A's projected yield, for each of the shuffle operations across 1000 unique random number seeds. As in Figure 5 below, each box extends from the upper to the lower quartile value, and each red line shows the median. The whiskers indicate the most extreme value within 1.5 * the inner quartile range, with values beyond this illustrated with a '+'. The distance between the unperturbed model's projected yield and observations is represented by the dotted lines. Only seasonal shuffling was performed on yield inputs, since this dataset is comprised of per-season values.

586

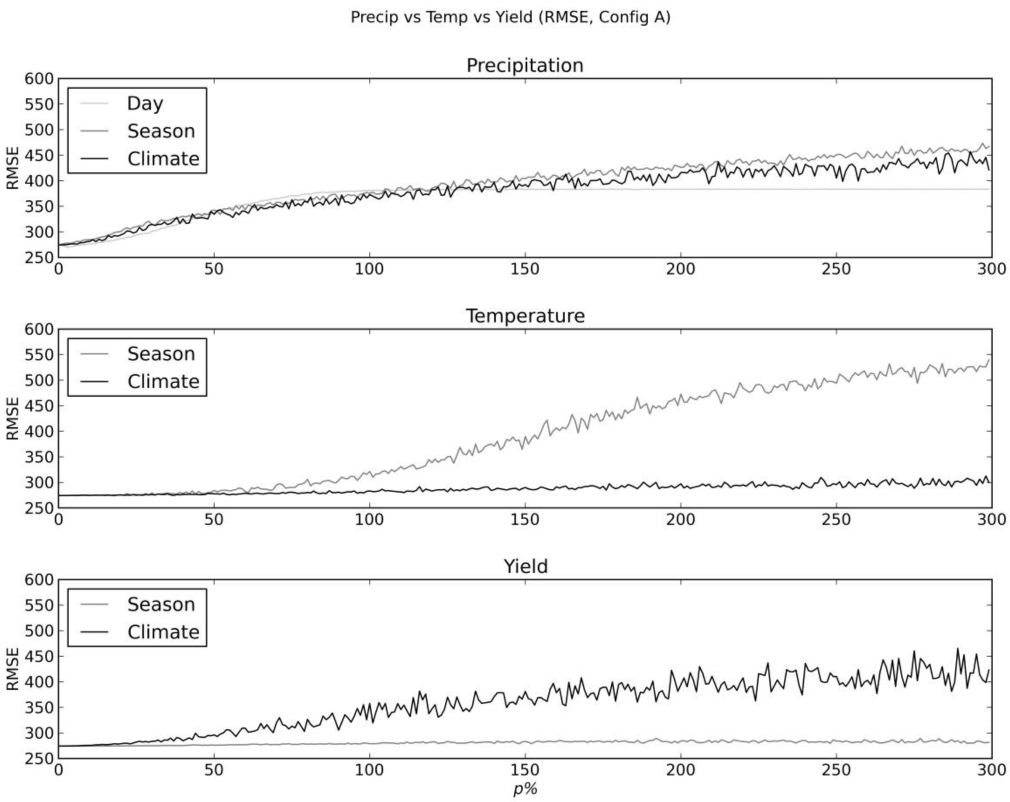

Figure 3 Mean RMSE of projected model yield compared to observed yield for increasing *p* (configuration A).

587

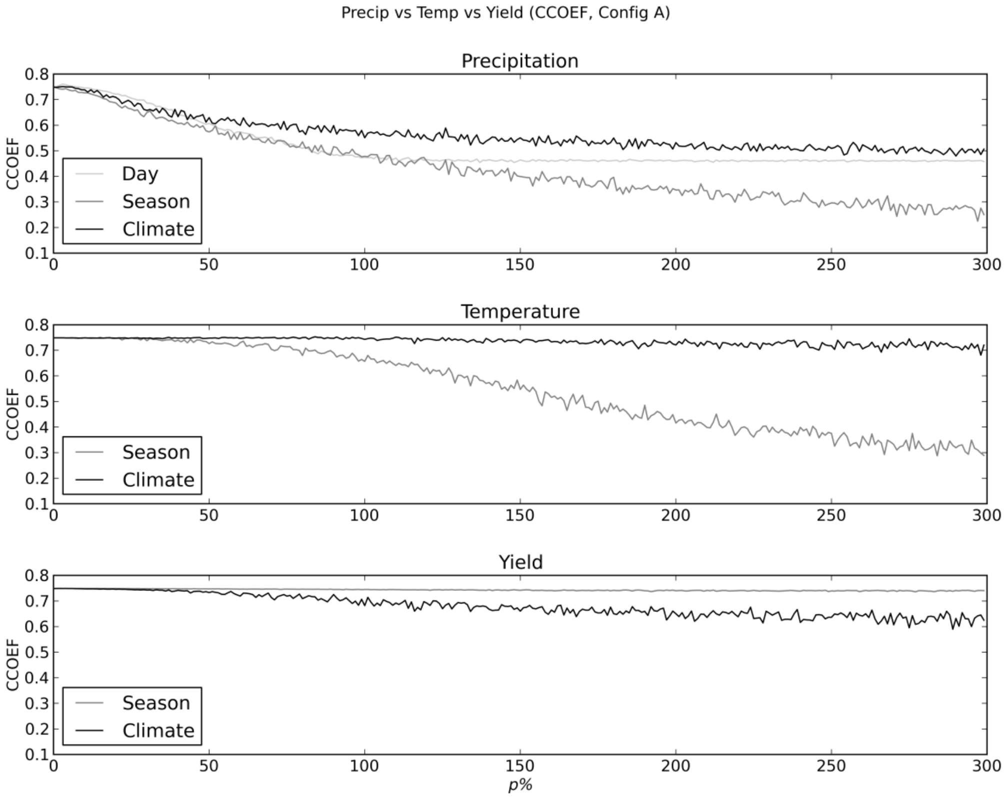

Figure 4 Mean correlation coefficient of projected model yield compared to observed yield as *p* is increased (configuration A).

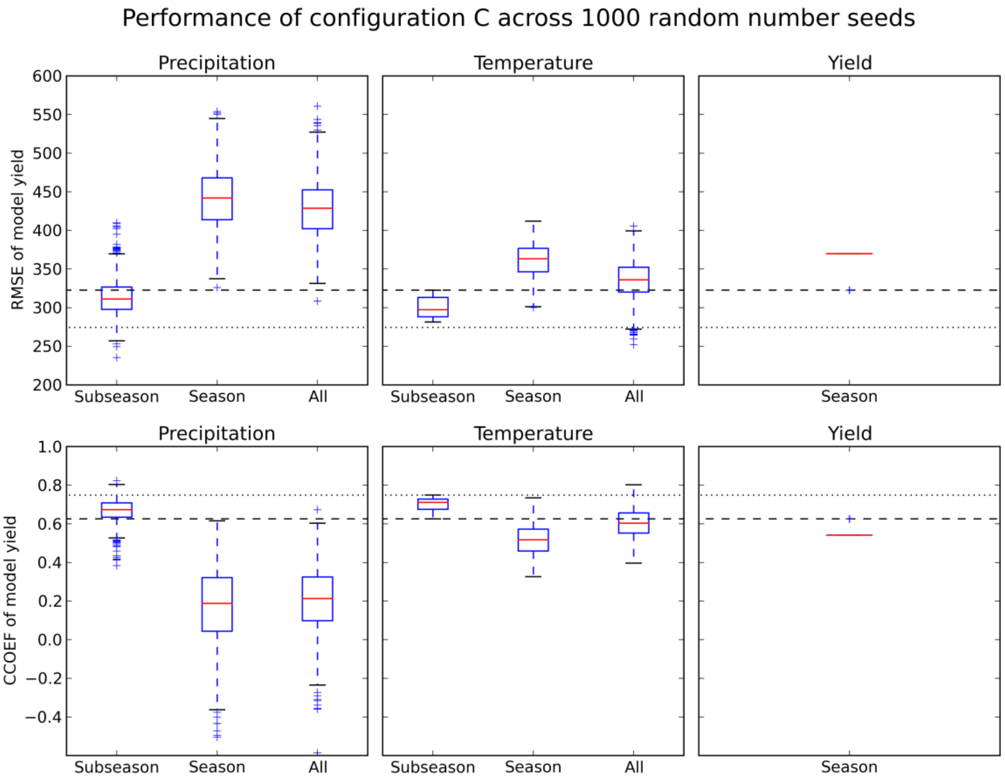

Figure 5 Performance of GLAM configuration C for each shuffle operation. As in Figure 2, the dotted lines represent the distance between observations and the unperturbed configuration A yield projection. The dashed lines represent the distance of configuration C.

589

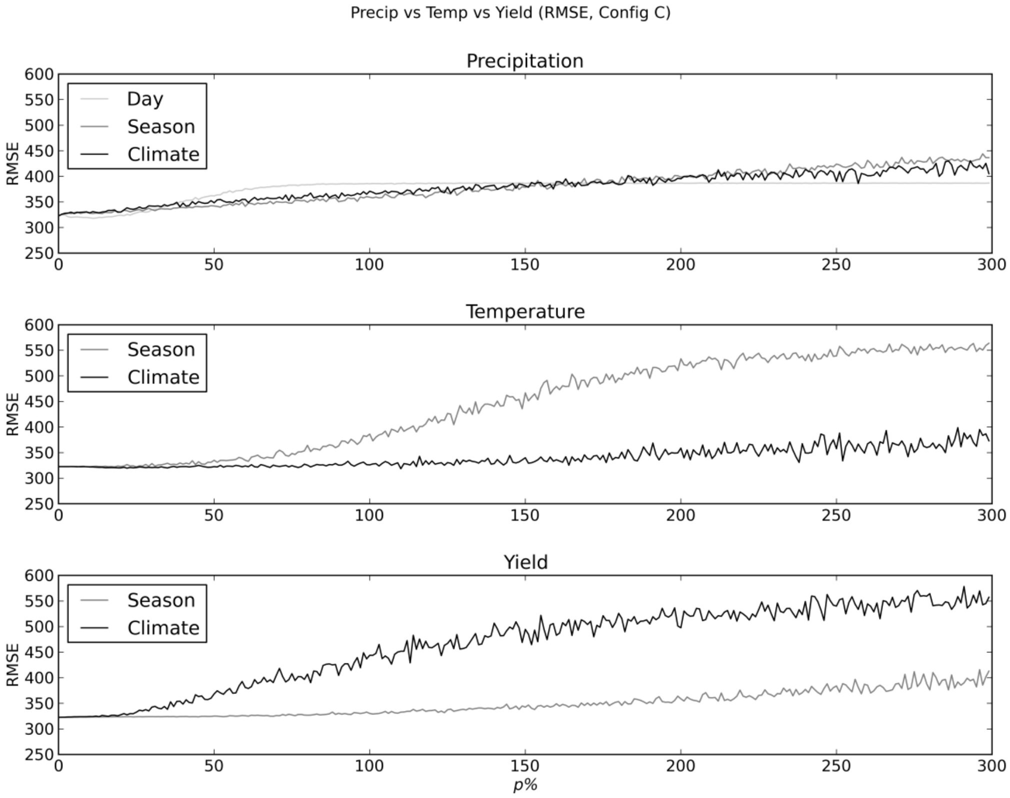

Figure 3 Mean RMSE of projected model yield compared to observed yield as *p* is increased (configuration C).

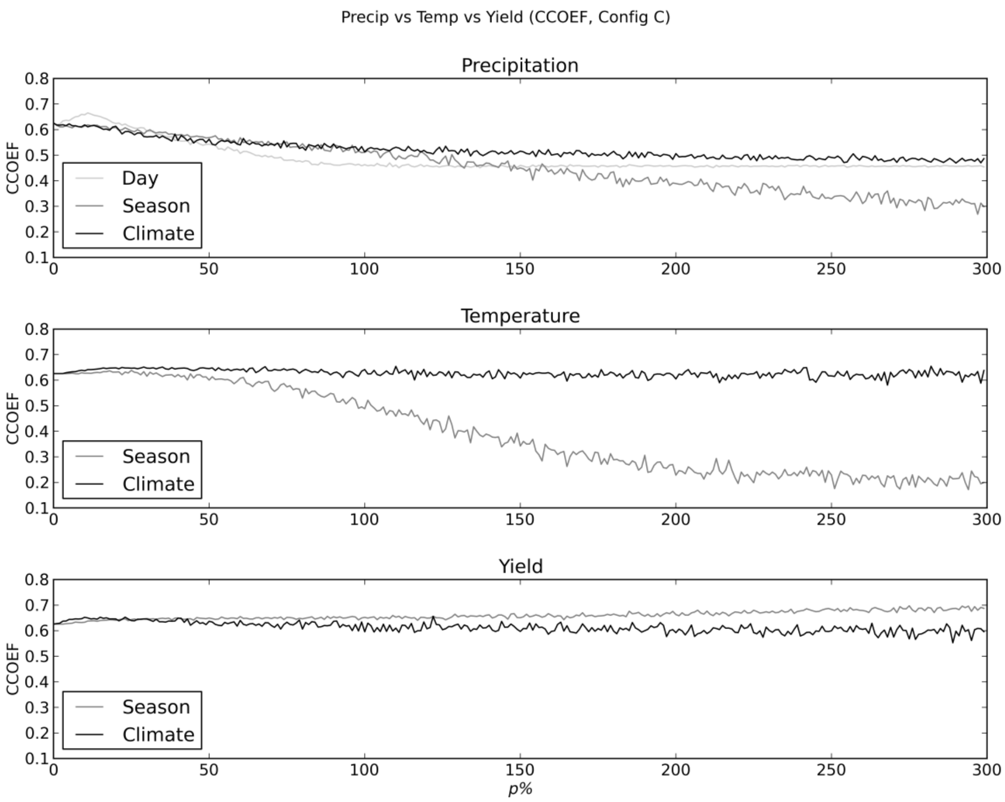

Figure 4 Mean correlation coefficient of projected model yield compared to observed yield as *p* is increased (configuration C).

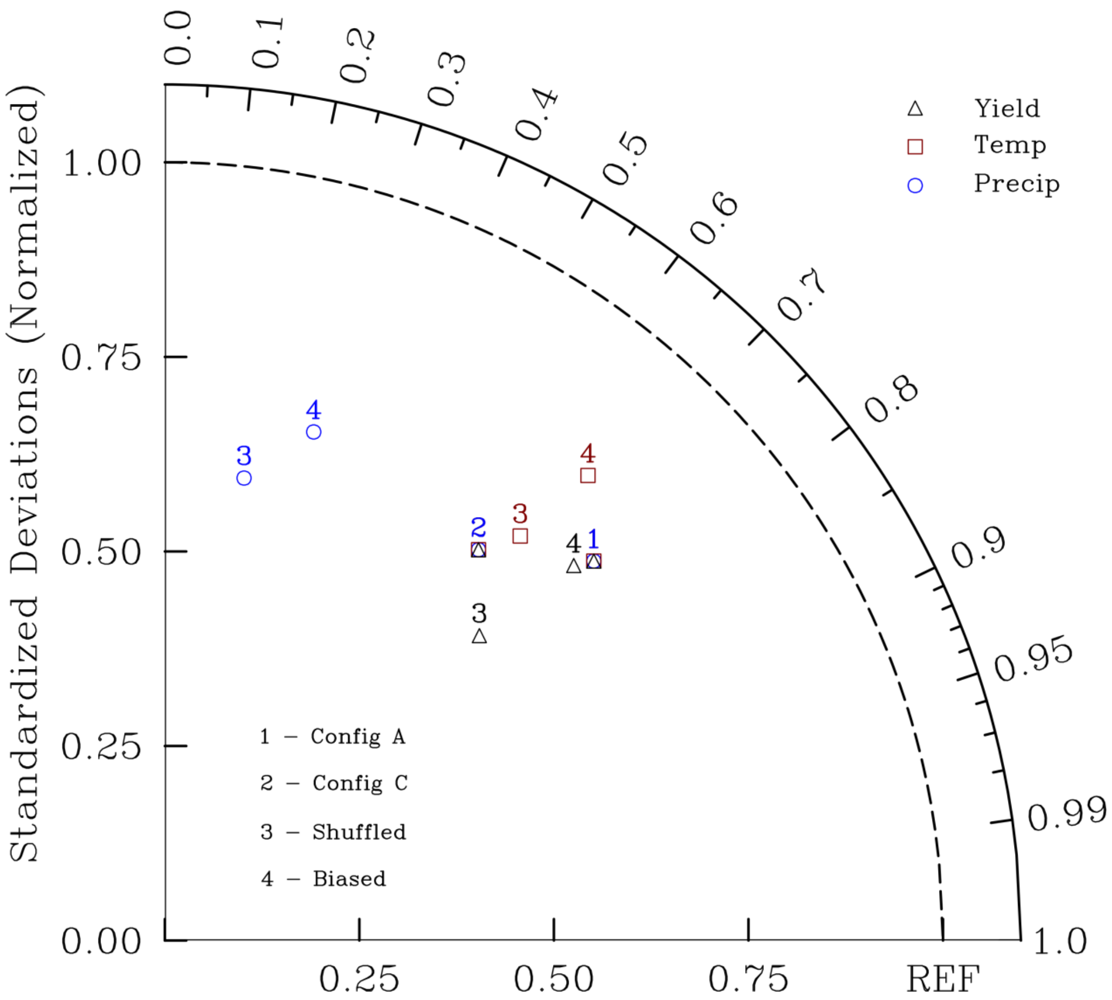

Figure 5 Comparison of the mean correlation coefficient and mean standard deviation (normalized to observations) of each data scheme, for configuration A at the seasonal timescale. For each Bias type, the single value of *p* whose mean RMSE was closest to the equivalent shuffled RMSE was chosen. The performance of the control runs of configurations A and C are also shown.

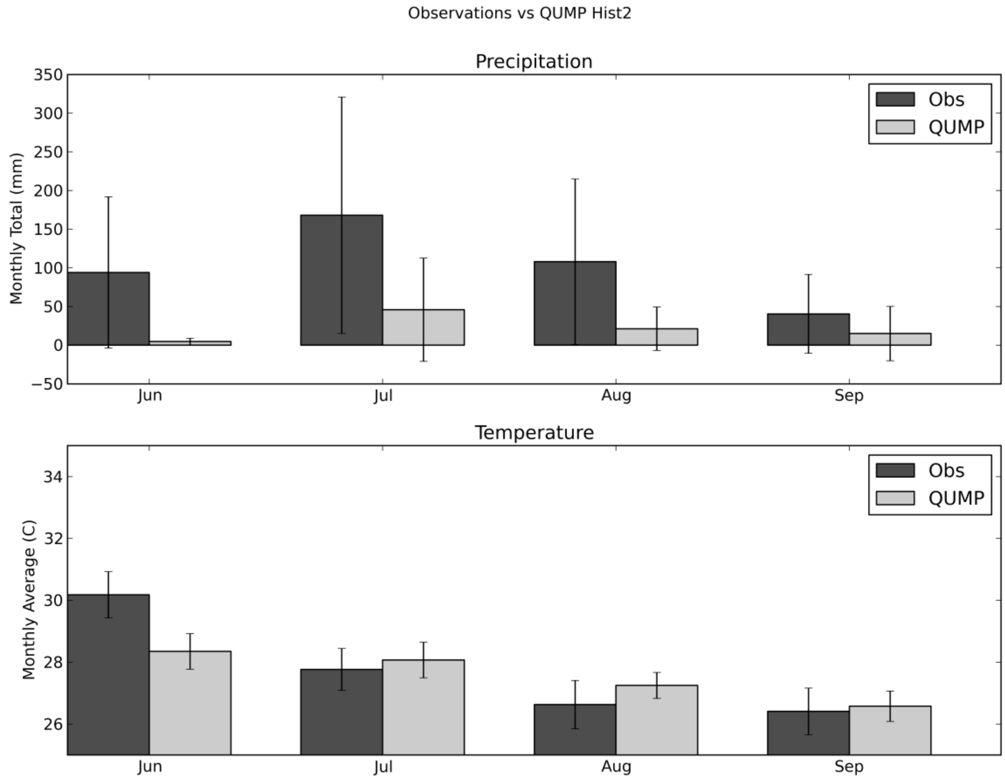

Figure 6 Comparison of monthly precipitation and temperature observations with the Hist2 control run of the QUMP 17-member HadCM3 ensemble. The mean and standard deviation for each month in the growing season is shown for the years 1966-1989.

593