Emulating global climate change impacts on crop yields

Emulating global climate change impacts on crop yields

Oluwole K. Oyebamiji [1], Neil R. Edwards [1], Philip B. Holden [1], Paul H. Garthwaite [2], Sibyll Schaphoff [3], and Dieter Gerten [3]

[1] Department of Environment, Earth & Ecosystems, The Open University, Milton Keynes, UK

[2] Department of Mathematics & Statistics, The Open University, Milton Keynes, UK

[3] Potsdam Institute for Climate Impact Research (PIK), Telegrafenberg A62, 14473 Potsdam, Germany

---

Address for correspondence: Oluwole K. Oyebamiji, Department of Environment, Earth & Ecosystems, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK.

E-mail: Oluwole.Oyebamiji@open.ac.uk.

Phone: (+44) 7411 875 750.

Fax: (+44) 1908 655 151.

Abstract: The potential effects of climate change on the environment and society are many. In order to effectively quantify the uncertainty associated with these effects, highly complex simulation models are run with detailed representations of ecosystem

processes. These models are computationally expensive and can involve a computer run of several days. Computationally cheaper models can be obtained from large ensembles of simulations using statistical emulation. The purpose of this paper is to construct a cheaper computational model (emulator) from simulations of the Lund-Potsdam-Jena managed Land (LPJmL), which is a dynamic global vegetation and crop model. This paper focuses on statistical emulation of potential crop yields from LPJmL and an emulator is constructed using a combination of ordinary least squares, principal component analysis and weighted least squares methods. For five climate models, under cross-validation the percentage of variance explained ranges from 60-88% for the rainfed crops and 62-93% for the irrigated crops. The emulator can be used to predict potential crop yield change under any future climate scenarios and management options.

Key words: Crop yield; CO_2 fertilization effect; principal components; distance weighted regression

1 Introduction

The world population is projected to increase by 35% by the middle of this century (UNFPA, 2010). This will cause a rise in demand for major food crops that will require a considerable increase in crop production. Climate change, food insecurity and how to effectively feed over 9 billion people by mid-century are major problems threatening humanity (Smith and Gregory, 2013). Rising temperatures and CO_2 levels and changing precipitation patterns will affect crop production (Parry *et al.*, 2004).

The IPCC Fifth Assessment Synthesis Report (Section SPM2.3) notes that "Global temperature increases of 4° C or more above late 20-th century levels, combined with increasing food demand, would pose large risks to food security globally (high confidence)." The problem requires multi-disciplinary approaches. A robust and coherent assessment of the climatic impact on future crop-yields is essential to inform policy makers. Quantification of the reduction of climate change impacts in different sectors by moving from a no-mitigation approach to several alternative mitigation scenarios was the major focus of Warren *et al.* (2013) and further evaluation of uncertainty associated with impact reduction in Arnell *et al.* (2013). These studies concluded that urgent global measures can prevent the larger impacts of climate change that are otherwise projected to occur by mid-century.

The relationships between crop yields, weather and climate have attracted considerable attention. Many authors have applied empirical methods. Simulated and historical data were explored to assess the global climatic impact on crop yields in Lobell and Burke (2008, 2010). An earlier study by Kart (1979) examined the relationship between crop and weather using a ridge regression approach. Reddy and Pachepsky (2000) estimated changes in crop yield from monthly weather projections of climate variables. Wallach (2011) extended the Kart (1979) approach by estimating wheat production using multiple regression. Bornn and Zidek (2012) evaluated wheat yields using a Bayesian method, examining the significance of incorporating spatial information in crop-yield modelling. Schlenker and Roberts (2006) investigated the effect of change in average weather on crop-yield, focusing especially on the non-linear effect of temperature on growing season. Matis *et al.* (1989) applied a non-parametric Markov chain approach to crop yield prediction while Kim *et al.* (2005) derived a Bayesian bootstrap method to derive the posterior distribution of

the parameters rather than the distribution of a sample statistic. Lobell *et al.* (2006) developed a non-linear model to relate crop-yields to weather and climate, finding a significant non-linear relationship between temperature and yields. These statistical approaches were all applied at relatively small spatial scales (typically country level or smaller), greatly simplifying the problem compared to the global scale analysis we perform here. Bornn and Zidek (2012) note that forecasting wheat yields across the Canadian prairies is in itself a challenging task due to substantial spatial variability.

The other general approach is the application of process-based models (Bondeau *et al.*, 2007; Fader *et al.*, 2010) that simulate detailed physical and biological processes. Leemans and Solomon (1993) implemented a water balance model within a geographic information system (GIS) that integrates databases, while Fischer *et al.* (2001) extended the approach with the Global Agro-Ecological Zones (GAEZ) software. Similarly, Rosenzweig *et al.* (2014) used a process-based approach to evaluate the global consequences of various climate change scenarios on crop productivity. Recently, Muller and Robertson (2014) compared the performance of two global crop models - the LPJmL of Bondeau *et al.* (2007) and Decision Support System for Agrotechnology Transfer (DSSAT) of Jones *et al.* (2003).

Future projections encompass many uncertainties. These include uncertain future scenarios for greenhouse gases and other emissions, uncertain climate projections for a given scenario, with different global climate models (GCMs) projecting different climates for the same scenario, and uncertain crop yields in response to a given climate change projection. A significant source of crop yield uncertainty results from the poorly quantified "CO_2 fertilization effect", the observed increase in photosynthesis rates that arises when ambient CO_2 concentrations are increased. Impact studies are

generally forced to address a subset of these uncertainties. For instance Lobell and Burke (2010) focussed on just one crop from Sub-Saharan Africa while Osborne and Wheeler (2013) assessed the impact at a coarse scale of country level. In a broader study, Osborne *et al.* (2013) determined projections from fourteen different GCMs but considered only one emission scenario and one time point. Deryng *et al.* (2011) used two GCMs and two adaptation scenarios but concentrated only on rainfed crops and neglected the CO_2 fertilization effect.

Here we integrate the empirical and process-based approaches with the statistical emulation of an ensemble of simulations of the process-based LPJmL model. Computational speed is highly problematic for coupling complex models together. Emulation is a tool for simplification of models that leads to reduced-form representations of complex models that are computationally much faster and hence easier to couple to other models. Emulation offers a rapid alternative for the projection of crop productivity under diverse climate scenarios, but at the same time it allows us to capture important relationships between crop yield and climate, enabling realistic prediction of responses to climatic change. Emulation would further facilitate other calculations (e.g. sensitivity analysis) that would not be practical using the LPJmL model directly.

The approach uses a combination of OLS, PCA and WLS regression to model the potential crop yields that are simulated by LPJmL. We use a two-stage method. The first stage applies a least squares regression analysis similar to Holden *et al.* (2010) to model change in potential crop yields as a function of climate change and other relevant covariates. The second-stage uses a novel combination of PCA with a WLS regression for interpolation of residual variation that is left unexplained by

the regression equation. This has similarities to work of Higdon *et al.* (2008), who combine GP emulation with a basis representation (such as principal components) for calibration of computer models with high dimensional output. Our approach is also related to the technique suggested in O'Hagan (2006), which involves combining a regression with a GP emulation of the regression's residuals for improving emulator performance.

The emulators provide high-resolution spatial output fields for five major crop types. Emulated projections are provided for both irrigated and rainfed crops, and for variable degrees of crop management intensity. They also address the uncertainty due to the CO_2 fertilisation effect.

In Section 2 we describe the models and simulation set up. In Section 3 we describe the methods and implementation of the emulator algorithm. Section 4 reports the results of the emulations. Section 5 discusses the results and gives concluding comments. Supplementary material to accompany this paper is available at http://??? (will link to an address in the journal's repository.)

2 Models used for the analysis

We built the crop-yield emulators using simulation output from the following models: LPJmL, Model for the Assessment of Greenhouse Gas Induced Climate Change (MAGICC) and Spatial Climate Generator (ClimGen). MAGICC is a simple carbon cycle climate model that simulates greenhouse gas (GHG) cycles, radiative forcing, and ice melt. The gas cycle uses standard formulae to convert surface emissions of gases to atmospheric concentrations and these, in turn, are then converted to radia-

tive forcing. The generated radiative forcing is then used to drive a diffusive energy balancing model to estimate global climate change. MAGICC6 is a new version of MAGICC (Meinshausen *et al.*, 2011) and is able to simulate global mean temperature (GMT) trajectories based on the emulation of the 18 Atmospheric Ocean General Circulation Models (AOGCMs) used in Solomon (2007) for the Fourth Intergovernmental Panel on Climate Change (IPCC) assessment report.

ClimGen is a spatial climate scenario generator. Emulated global mean temperature trajectories of a particular GCM from MAGICC6 are used to drive the ClimGen model. ClimGen uses a pattern-scaling method and produces spatial climate change information for a given global-mean temperature change. The method is based on the assumption that the pattern of climate change simulated by the coupled AOGCMs is relatively constant but the amplitude changes. These normalised patterns of climate change usually show considerable variation between different AOGCMs, and it is this variation that ClimGen is designed to explore (Osborn, 2009).

The LPJmL model of Bondeau *et al.* (2007) is a dynamic global vegetation model (Sitch *et al.*, 2003) enhanced to represent global agriculture in addition to natural vegetation. It uses eco-physiological relations and plant trait parameters for the estimation of photosynthesis, plant growth, maintenance and regeneration loss, fire disturbance, soil moisture, runoff, evapo-transpiration, irrigation and vegetation structure. Agricultural productivity is simulated through varieties of crop functional types (CFTs), both rainfed and irrigated. The LPJmL model takes as inputs climate variables such as precipitation, temperature and insolation, which are then disaggregated to quasi-daily values by a weather generator. The monthly input and output data are spatially explicit time series of about 60 000 global 0.5° resolution grid cells. Each

grid cell can contain a variety of natural or agricultural vegetation types, whose daily growth and productivity is simulated. Bondeau *et al.* (2007) described further the performance of the LPJmL simulation in term of crop yields, crop phenology and carbon fluxes.

In summary, MAGICC6 generates and provides future trajectories of global mean temperature that are used by ClimGen. ClimGen emulates the spatial response patterns by disaggregating the temperature trajectories to 0.5° spatial resolution of climate change patterns for temperature, precipitation and cloud cover. The generated climate scenarios are supplied as inputs to run the LPJmL simulation for the assessment of climate impacts on variables such as potential crop yield.

2.1 LPJmL simulation

LPJmL was run on seven climate change patterns, namely, Canadian Centre for Climate Modelling and Analysis Coupled Global Climate Model (CCCMA-CGCM31), Center for Climate System Research and Model for Interdisciplinary Research on Climate (CCSR-MIROC32HI), CCSR-MIROC32MED and Hadley Centre Global Environmental Model, Met Office United Kingdom (UKMO-HADGEM1), Goddard Institute for Space Studies (GISS-MODELEH), GISS-MODELER, and Institut Pierre Simon Laplace (IPSL-CM4) generated using ClimGen, which used trajectories of global mean temperature constructed by MAGICC6. In the MAGICC6 model, the forcing pathways of all four Representative Concentration Pathways (RCPs) were used. These RCPs are widely used in climate research. They describe alternative possible future emission scenarios and are designed to standardise climate model simulations for inter-comparison. The pathways are characterised by either long-term

stabilisation within the period $2100-2150$ (RCPs 4.5 and 6.0), progressively increasing forcing by 2100 (RCP 8.5), or decreasing forcing by 2100 and beyond (RCP2.6, also named RCP-3PD) (Moss *et al.*, 2010).

The simulations involve a spin-up stage that was used to equilibrate the long-term carbon stores (in natural and agricultural ecosystems) by repeating the observed climate ($1901-1930$ period) 33 times, immediately followed by an additional 13 repetitions in order to incorporate landuse change for reconstruction of historical soil carbon pools (Fader *et al.*, 2010). Then simulations followed for the period $1931-2000$, with transient climate and land use change data. The scenario period ($2001-2100$) for the different RCPs and GCM patterns started from year 2001. Land use change pattern and irrigation were held constant at their year 2000 values. Changes in the potential crop yields were simulated for each of five crop types under both rainfed and irrigated conditions. Results were calculated on a global $0.5^\circ \times 0.5^\circ$ degrees resolution.

3 Methods

3.1 A procedure for statistical emulation

In addition to the simulations described in 2.1, we also examined seven different crop management levels for each scenario, and performed simulations with and without the CO_2 fertilization effect. The CO_2 fertilization effect is the rise in crop yield as a result of elevated CO_2 in the atmosphere. In calibrating crop management, the Leaf Area Index (LAI) is a key parameter. LAI is the ratio of total upper leaf

surface of vegetation divided by the surface area of the land on which it grows. Crop management levels are represented by maximum leaf area index LAI_{max}, which in LPJmL represents a proxy for vegetation density (thus reflecting the vegetation response to the overall management intensity). Together with other synchronously varied parameters it is used to calibrate the modelled yields with respect to observed yields that are a function of local management practice, as in Fader *et al.* (2010). Here we use seven simulations with fixed parameters for each grid cell and crop type so as to derive the yield levels that would be achieved if those management levels were in place.

The LPJmL crop-yield data were based on simulations for 59199 grid cells on 0.5° by 0.5°, but here we only consider those cells where crops are actually simulated to grow (the other cells are coded zero). We built separate emulators for the rainfed and irrigated crops. The following five CFTs were selected for emulation: temperate cereal, rice, maize, groundnut and an oil crop, which is the maximum yield among soybean, sunflower and rapeseed. The five crops are chosen because they are widely grown across the globe. Cereal, rice and maize are very prominent staple food crops and provide over 50% of all calories consumed by the world. Oil is used domestically as a vegetable oil and also forms ingredients used in manufacturing products.

The average decadal yield given by LPJmL from 2005-2095 was computed for each crop in each grid cell. We then obtained the change in yields relative to the baseline decadal average in 2005-2014. We calculated the change in seasonal climate variables for the 37 input variables listed in Table 1. The emulators were constructed in two stages. The first stage uses an OLS regression method to fit a linear model that explains much of the variation between the response and input variables. The second

Table 1: The emulator's input variables.

Variables	Full names
scld/weld/spcld/acld	Change in mean cloud cover in summer/winter/spring/autumn
spre/wpre/sppre/apre	Change in mean precipitation in summer/winter/spring/autumn
stmp/wtmp/sptmp/atmp	Change in mean temperature in summer/winter/spring/autumn
swet/wwet/spwet/awet	Mean change in wet day frequency in summer/winter/spring/autumn
iscld/iweld/ispcld/iacld	Initial (baseline) mean cloud cover in summer/winter/spring/autumn
ispre/iwpre/isppre/iapre	Initial mean precipitation in summer/winter/spring/autumn
istmp/iwtmp/isptmp/iatmp	Initial mean temperature in summer/winter/spring/autumn
iswet/iwwet/ispwet/iawet	Initial wet day frequency in summer/winter/spring/autumn
co2 and cco2	Baseline mean CO_2 and change in mean CO_2
soil, lat and LAI	Soil and latitude parameters, and maximum leaf area index

In the northern hemisphere, *summer* = {June July August}, *winter* = {December January February}, *spring* = {March April May} and *autumn* = {September October November}; obvious changes are made for the southern hemisphere.

stage combines PCA with a WLS regression to explain some of the residual variation that is left unexplained by the first stage. We built a single emulator for the two CO_2 fertilization levels ("on" and "off"), but treat irrigated and rainfed crops separately. This allows the emulator to be flexible in predicting yield changes for any level of CO_2. Figure 1 shows the average yields of cereal (upper plots) and rice (lower plots) in the baseline period (2005-2014) given by LPJmL. The plots are for a moderate pathway (RCP6), the climate model CCSR-MIROC32HI, with CO_2 fertilization 'on' and management level 6. The crop is not grown in areas in white, such as most of America, Africa and Australia for irrigated rice. Corresponding plots for maize, oil and groundnut are given in Supplementary material S1.

The emulators were built from two GCMs, CCCMA-CGCM31 and CCSR-MIROC32HI,

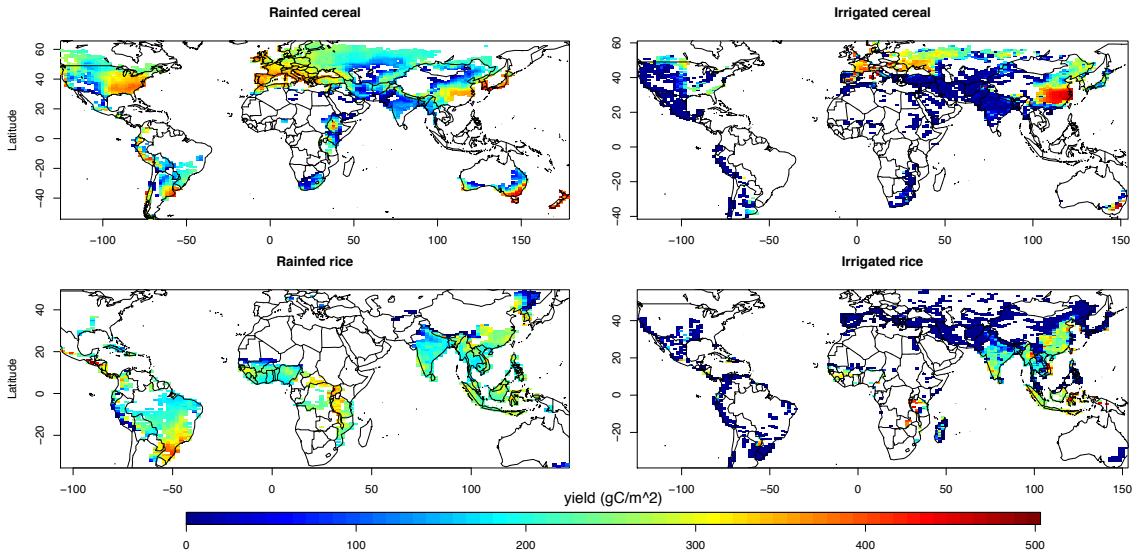

Figure 1: LPJmL average annual yield of cereal (upper two plots) and rice (lower two plots) in 2005-2014 for RCP6, CCSR-MIROC32HI, with CO_2 fertilization and management level 5. Left-hand plots: rainfed crops; irrigated crops. Areas in white over land correspond to grid cells with observations of zero.

four RCPs, and two simulation categories (with and without CO_2 fertilization effect), giving 16 ($2 \times 4 \times 2$) different scenarios. Each scenario has seven crop management levels and eight time-slices, with each time-slice consisting of 59199 observations. After removing the zero observations from the data, there are $16649, 6913, 19139, 7100$ and 8427 valid grid points per scenario for rainfed temperate cereal, rice, maize, groundnut and oil, respectively, and 8963, 7325, 9146, 2386 and 2731 for irrigated crops. In order to evaluate the performance of the LPJmL emulators another five GCMs, UKMO-HADGEM1, GISS-MODELER, GISS-MODELEH, IPSL-CM4 and CCSR-MIROC32MED were used for cross-validation purposes.

We used all valid grid points for irrigated oil and groundnut emulators because each of these crops has less than 5000 non-zero observations in each scenario. For other crops, in the first stage we used observations from a random sample of 5000 valid grid points because the stepwise algorithm could not fit the whole dataset. The 5000 grid points are fixed across the time-slice, RCP, GCM and management levels as well as simulation categories. We sampled about 4.5 million observations on each input variable. We then fitted a single model to the sampled data for each crop-yield when crops are rainfed. This procedure was carried out for each of the five crops and repeated for the case where crops are irrigated. For the second stage, all valid grid points were used.

Two climate models were used in the construction of the emulators so as to better span the possible climate input space. We should note though, that using more climate models only help spans 'climate model space', not necessarily 'true climate input space'.

However, we also used several RCPs, which incorporates information on a broad range

of emission pathways, and a minimum of 2386 grid points for each emulator. Hence, the climate values that occur across grid points in our training data cover a very wide range. This helps our emulators predict the change in crop yield given by LPJmL for a range of climate forcing scenarios that are not restricted to the RCPs and climate models (GCMs) used to construct them: the climate states that arise within each grid point should resemble the climate that has arisen somewhere in the training data. In section 4.1 we use cross-validation to examine emulator performance, testing the emulators on five climate models that played no part in their construction.

3.1.1 Stage 1

An emulator is constructed in two stages. In stage 1 stepwise regression is used to obtain a parsimonious model for predicting crop yield change (relative to baseline) from climate and other explanatory variables. The response variable is the change in yield given by LPJmL and is denoted by y. As noted earlier, each combination of RCP, GCM and CO_2 fertilization level is referred to as a scenario, giving 16 scenarios. Let n denote the number of grid cells being used for the combination of crop and irrigation regime of current interest. (So n equals 2386 or 2731 when emulating irrigated oil or groundnut, respectively, and 5000 for other emulators.) For each scenario we have $7 \times 8 \times n$ data as LPJmL gave values for seven management levels and eight different time slices. Hence y has $16 \times 7 \times 8 \times n$ data values. The explanatory variables are listed in Table 1.

An integer with values between 1 and 7 was used to represent the LAI_{max} parameter and this formed a factor variable in the regression analysis. The other explanatory variables can enter the regression as linear or quadratic terms. All two-way interac-

tions were also considered for inclusion. Thus spre, spre^2 and spre.wwet are examples of the potential terms in the regression model.

The emulators were built using the Revolution *R* Enterprise, which has a mechanism for scaling data to handle big computations. We used a bi-directional stepwise regression that performs variable selection by combining both forward selection and backward elimination. In the forward stepwise mode, a linear model is fitted that starts from a null model and variables are included step by step until the Akaike Information Criterion (AIC) judges that the algorithm has converged. This process is followed by backward elimination where non-significant variables are removed step by step until the Bayesian Information Criterion (BIC) determines that no further variables should be removed. AIC is a more lenient criterion than BIC so a long model is built and then simplified.

3.1.2 Stage 2

In contrast to stage 1, here we formed a separate emulator for each combination of management level and time slice. Also, all grid-points with non-zero observations are used, whereas in stage 1 we generally took a sample so as to cap their number at 5000. Rather than having multiple subscripts to indicate crop, irrigation regime, management level and time slice, we will consider just a single crop/irrigation regime/management level/time slice combination and let $\mathbf{y}_j$ be the vector of changes in yield given by LPJmL for that combination in the j^{th} scenario ($j = 1, \ldots, 16$). We let $\tilde{\mathbf{y}}_j$ be the corresponding predictions given by the stage 1 emulator and $\boldsymbol{\varepsilon}_j = \mathbf{y}_j - \tilde{\mathbf{y}}_j$ is the error in prediction. Each $\tilde{\mathbf{y}}_j$ and $\boldsymbol{\varepsilon}_j$ is an $N \times 1$ vector, where N denotes the number of grid cells with non-zero observations for that crop/irrigation regime. As

$\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{16}$ are predictions on *training* data, the values of $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{16}$ are known.

Given a new vector of predictions, $\tilde{\mathbf{y}}^\star$, from a fresh scenario where the LPJmL values are unknown, the aim is to estimate the error of $\tilde{\mathbf{y}}^\star$ from $\mathbf{E} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{16})$ and $\tilde{\mathbf{Y}} = (\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{16})$. We apply a PC decomposition to the $16 \times N$ matrix $\tilde{\mathbf{Y}}^{\mathrm{T}}$ and select just the first four principal components.

The resulting 16×4 matrix $\mathbf{X}_0$ of PC scores given by these four components is then used as explanatory variables for the WLS regression of our residual patterns $\mathbf{E}$. Details are given in Appendix 1.

Our method in this stage is similar to a pattern scaling approach that is commonly used in climate scenario generation. Pattern scaling assumes that, given any particular point in space and time, there exists a linear relationship between climate change pattern and global mean temperature with a constant spatial pattern. Here, we allow for a more general multilinear relationship in the residual (Holden et al., 2014). The residual patterns from the OLS results in stage 1 indicated that the residual patterns are relatively similar across RCP and GCM (see Supplementary Materials S2). Hence, for example, if a grid cell has a negative residual in one scenario, then that grid cell is likely to have a negative residual for other scenarios. Stage 2 exploits the similarity between the error patterns across scenarios.

Having obtained the residuals $\mathbf{E}$ by calculating the differences between the emulator predictions and the actual LPJmL values for each scenario, we then interpolate these residual patterns using distance-weighted regression. More weight is assigned to known scenarios that are similar in pattern to the unknown scenario, with similarity determined by the distances from the known scenario points $\tilde{\mathbf{y}}_i, \ldots, \tilde{\mathbf{y}}_{16}$ to the

unknown scenario $\tilde{\mathbf{y}}^\star$; closer scenarios are more similar and receive greater weight.

Several distance metrics were tried: linear, square, cubic, spherical, power exponential, Gaussian and three forms of Matern metric. Details of these metrics are given in Appendix 2 where a variogram is also given that illustrates the need to take account of the distance between scenario points when modelling stage 1 prediction error. Performances of the metrics under validation data were compared. We chose a squared distance method scaled by their eigenvalues because it is amongst the best metrics in terms of the proportion of variance it explained and it is the simplest of the better methods. The weights are chosen to be inversely proportional to the squared distance and the same weighting scheme is applied across all grid points. The weights form the non-zero elements of a 16×16 diagonal matrix, $\mathbf{W}$.

A separate weighted regression is performed for each grid cell. For the i^{th} cell, the dependent variable is the i^{th} row of $\mathbf{E}$, $(i = 1, \ldots, N)$ but the data matrix for the explanatory variables ($\mathbf{X}_0$) and the weight matrix ($\mathbf{W}$) are the same for each grid cell. Thus most of the calculations for estimating regression coefficients need only be performed once, rather than once for each grid cell. From the stage 1 predictions for the new scenario, $\tilde{\mathbf{y}}^\star$, we calculate its PC score ($\tilde{\mathbf{x}}^\star$) and use the regression equation for the i^{th} grid cell to estimate the error in prediction for that cell. If $\hat{\varepsilon}_i^\star$ is the resulting estimate then we revise the prediction for the cell by adding $\hat{\varepsilon}_i^\star$ to it.

The following is a summary of the calculations in Stage 2 (see Appendix 1 for details).

(i) Perform a principal components analysis of $\tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{Y}}$. The non-zero eigenvalues are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{16}$ and the corresponding eigenvectors are $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{16}$. Put $\tilde{\mathbf{x}}^\star = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{16})^{\mathrm{T}}\tilde{\mathbf{y}}^\star$ and $\tilde{\mathbf{x}}_j = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{16})^{\mathrm{T}}\tilde{\mathbf{y}}_j$ for $j = 1, \ldots, 16$.

(ii) Denote the k^{th} components of $\tilde{\mathbf{x}}^{\star}$ and $\tilde{\mathbf{x}}_j$ by $x_k^{\star}$ and x_{jk}, respectively. Then $w_1, \ldots, w_{16}$ are the non-zero elements of the diagonal matrix $\mathbf{W}$, where $w_j = (1/d_j^2)/\{\sum_1^{16}(1/d_j^2)\}$, with $d_j^2 = \sum_{k=1}^{16} \lambda_k (x_k^{\star} - x_{jk})^2$.

(iii) The explanatory variables for the WLS regression are constructed from the first four eigenvectors of $\mathbf{\Gamma}$. We put $\hat{\mathbf{\Gamma}} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \boldsymbol{\gamma}_4)$ and $\mathbf{X}_0 = \tilde{\mathbf{Y}}^{\mathrm{T}} \hat{\mathbf{\Gamma}}$.

(iv) Weighted least squares gives $\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_0^{\mathrm{T}} \mathbf{W} \mathbf{X}_0)^{-1} \mathbf{X}_0^{\mathrm{T}} \mathbf{W} \tilde{\boldsymbol{\varepsilon}}_i$ as the vector of regression coefficients for the i^{th} grid cell, where $\tilde{\boldsymbol{\varepsilon}}_i^{\mathrm{T}}$ is the i^{th} row of $\mathbf{E}$.

(v) The estimated error for the i^{th} grid cell is $\hat{\varepsilon}_i^{\star} = \hat{\boldsymbol{\beta}}_i^{\mathrm{T}} \mathbf{x}_{\mathbf{0}}^{\star}$, where $\mathbf{x}_0^{\star} = \hat{\mathbf{\Gamma}}^{\mathrm{T}} \tilde{\mathbf{y}}^{\star}$.

(vi) In order to avoid unusual values and prevent too much extrapolation from known residuals, we compare $\hat{\varepsilon}_i^{\star}$, with each component of $\tilde{\boldsymbol{\varepsilon}}_i$. Let $(\varepsilon_i^{min}, \varepsilon_i^{max})$ denote the range of these components. If $\hat{\varepsilon}_i^{\star}$ is outside this range, we set it equal to the range's nearer endpoint. Thus the revised estimate of y for the i^{th} grid point is $\tilde{y}_i^{\star} + \hat{\varepsilon}_i^{\#}$, where $\tilde{y}_i^{\star}$ is the i^{th} component of $\tilde{\mathbf{y}}^{\star}$ and

$$\hat{\varepsilon}_i^{\#} = \begin{cases} \varepsilon_i^{min} & \text{if } \hat{\varepsilon}_i^{\star} < \varepsilon_i^{min} \\ \varepsilon_i^{\star} & \text{if } \varepsilon_i^{min} \leq \hat{\varepsilon}_i^{\star} \leq \varepsilon_i^{max} \\ \varepsilon_i^{max} & \text{if } \hat{\varepsilon}_i^{\star} > \varepsilon_i^{max}. \end{cases} \tag{3.1}$$

Our methods described above are simple and flexible to apply. Diagnostic plots of the residuals from WLS regressions are reported in Supplementary Materials S3.

A Gaussian process model could not be applied directly to our data because of the computational difficulty from the large sample size coupled with the large number of parameters to be estimated. GP scales cubically with the number of observations $O(N^3)$, which is not appropriate for our present data – even after averaging decadally

and sampling from each scenario, the data matrix contains approximates 4.5 million values. It might be possible to use GP for residual interpolation, rather than WLS, but this would still have a high computational cost and it would be necessary to reduce the resolution and aggregate data to a country level in order to reduce the computational load.

3.2 Model performance

We assessed the performance of the emulator using five climate models that had not been used to construct the emulator, taking the proportion of variance (ρ) that the emulator explained as a measure of emulator efficiency. This proportion was calculated separately for each combination of climate model/crop/irrigation regime. For one combination, let $\bar{y}$ denote the average value given by LPJmL and let $\tilde{\mathbf{y}}^{\star}_{ijk\ell t}$ be the emulator final predictions for the i^{th} grid cell, CO_2 fertilization level j, management level k, RCP ℓ and time slice t. Also, let $y_{ijk\ell t}$ denote the actual LPJmL value to which the latter corresponds.

We compute the squared differences between the actual LPJmL values and $\bar{y}$ and also compute the squared differences between the LPJmL values and the emulator predictions. The proportion of the variance in the LPJmL values that is explained by the emulator is

$$\rho = 1 - \left[\frac{\sum_{t=1}^{8} \sum_{\ell=1}^{4} \sum_{k=1}^{7} \sum_{j=1}^{2} \sum_{i=1}^{N} (y_{ijk\ell t} - \tilde{\mathbf{y}}^{\star}_{ijk\ell t})^2}{\sum_{t=1}^{8} \sum_{\ell=1}^{4} \sum_{k=1}^{7} \sum_{j=1}^{2} \sum_{i=1}^{N} (y_{ijk\ell t} - \bar{y})^2} \right] \tag{3.2}$$

where, as before, N is the number of grid cells with non-zero observations for the

crop/irrigation regime of current interest. The overall cross-validation root mean squared error (RMSE_{CV}) is

$$\text{RMSE}_{CV} = \left(\sum_{t=1}^{8} \sum_{\ell=1}^{4} \sum_{k=1}^{7} \sum_{j=1}^{2} \sum_{i=1}^{N} \frac{(y_{ijk\ell t} - \tilde{\mathbf{y}}^{\star}_{ijk\ell t})^2}{(8 \times 4 \times 7 \times 2 \times N)} \right)^{1/2}. \tag{3.3}$$

3.3 Sensitivity analysis

Calculating the total effects of each explanatory variable helps identify the relative importance of variables in a model, that is, the contribution of each variable to the total variance. We use the Sobol global sensitivity method. This computes indices by decomposing the variance up to a specified order. The method we used computes the first order and total indices. Suppose our model is represented by $y = f(x_1, \ldots, x_p)$. The indices are given respectively as

$$S_j = \frac{Var[E(y|x_j)]}{Var(y)} \tag{3.4}$$

$$S_{Tj} = \frac{Var(y) - Var[E(y|x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p)]}{Var(y)} \tag{3.5}$$

where $Var[E(y|x_j)]$ is the first order or main effect of variable x_j, and $Var(y)$ is the total variance of the response y. S_{Tj} is the result of the main effect of x_j and all its interaction with other parameters up to order p (Saltelli, 2002).

4 Results

4.1 Emulator predictions

Figure 2 gives density plots over grid cells for the percentage change in crop yield (2084-2095) relative to the baseline yield for both CCSR-MIROC32HI and CCMA-

CGCM31 GCMs. We consider all management levels and a moderate emission scenario, RCP 6. The right-hand plots show the change in crop yield when there is no CO_2 fertilization while the left-hand plots show change in yield with CO_2 fertilization. In the right-hand plots, most crops show a preponderance of negative values, indicating a general reduction in yield. The exception is oil, which mainly shows positive values. The distributions each have a single major peak but vary as to whether they have a further minor peak (or even several minor peaks). The skewness of the density function also varies markedly with crop.

Now considering the left-hand plots (with CO_2 fertilization), the density plots show a preponderance of positive values and are more diverse in pattern, which could be a result of non-linear interactions between climate and CO_2. Comparison of the left- and right-hand plots shows that CO_2 fertilization has a marked effect on all crops except maize, for which the two plots are strikingly similar. (Maize is less affected by CO_2 fertilization because it is a C_4 plant and has a mechanism to efficiently transport CO_2 to the photosynthetic parts, limiting photorespiration rate thereby reducing water losses). Overall, the varying patterns in Figure 2 clearly show the diversity of the effects we are emulating. The changes in crop yields are characterized with high variability and there are varying patterns across different scenarios.

Figure 3 gives time series plots showing the percentage change in global crop yield over each decade relative to the baseline period under the CCSR-MIROC32HI, either without CO_2 fertilization (right-hand plots) or with CO_2 fertilization (left-hand plots). It shows temporal variability of the rainfed crops under the RCP6 scenario with a separate line for each management level. The sensitivity of change in yield to the CO_2 fertilization effect is apparent, with CO_2 fertilization greatly improving the change in

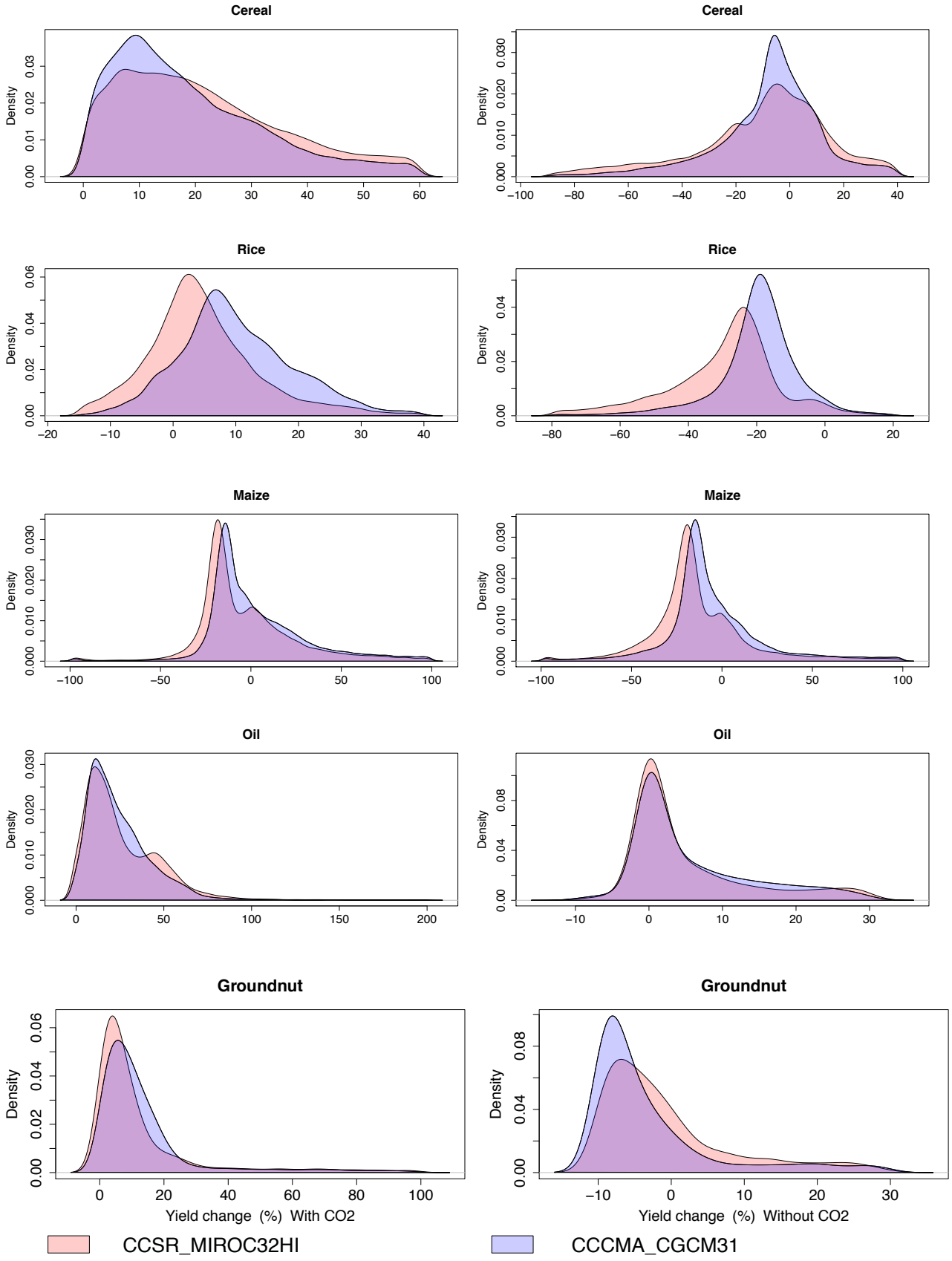

Figure 2: Probability distribution for the percentage decadal change between (2085-2094) and (2005-2014) for rainfed cereal, rice, maize, groundnut and oil respectively, RCP 6 and all management levels. Left-hand plots: with CO_2 fertilization; right-hand plots: without CO_2 fertilization.

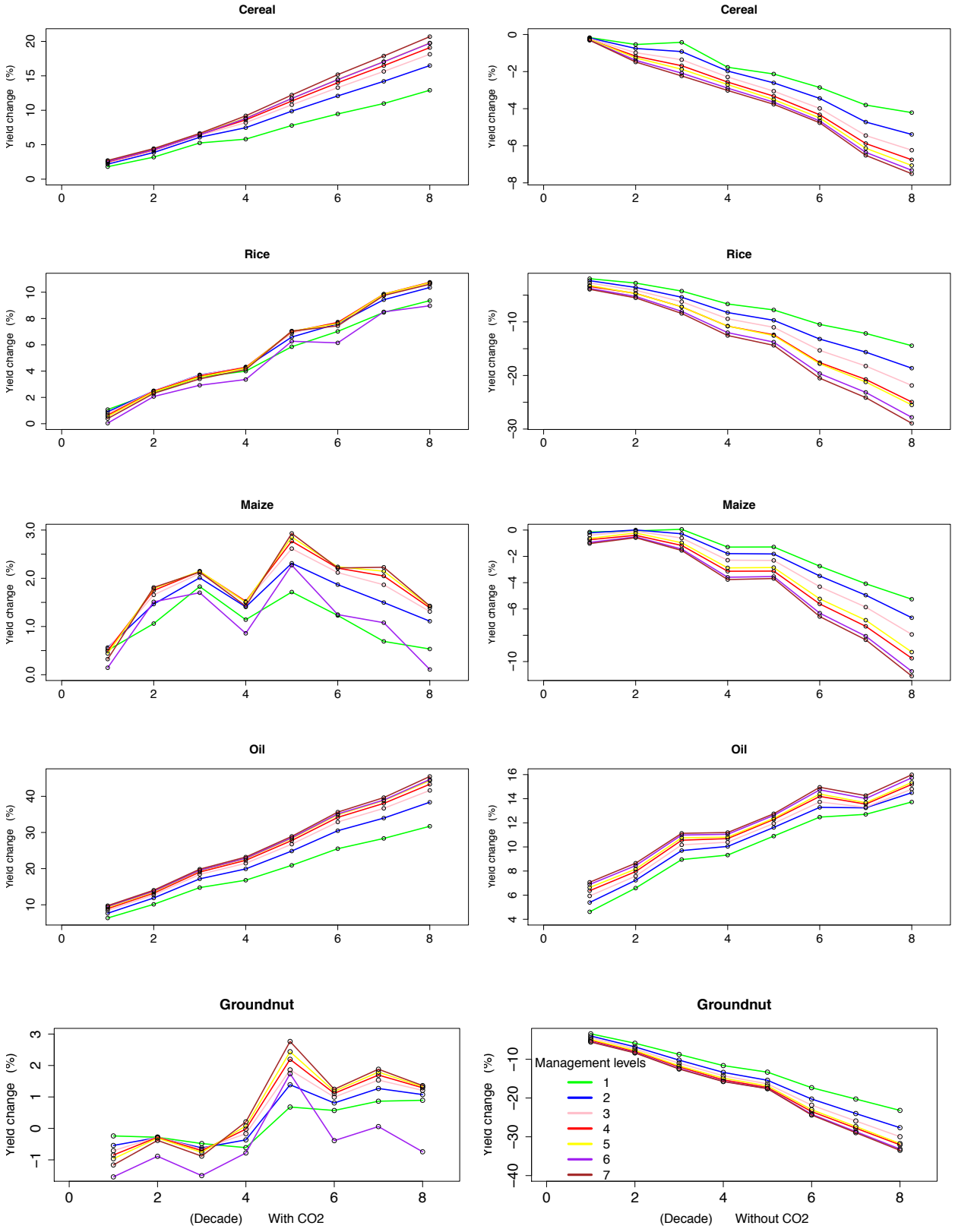

Figure 3: Time series plot showing the temporal pattern for the percentage decadal change relative to baseline period for rainfed temperate cereal, rice, maize, groundnut and oil respectively, for all time-slices, RCP 6 and all management levels for the CCSR-MIROC32HI average over all grid cells. Left-hand plots: with CO_2 fertilization; right-hand plots: without CO_2 fertilization.

Table 2: Cross-validated proportion of variance ρ and root mean squared error RMSE_{CV} showing the overall performance of the emulators for rainfed and irrigated crops, with all management levels, RCPs, and time slices, but with CO_2 fertilization only, for UKMO-HADGEM1.

Crop	1^{st} stage ρ		2^{nd} stage ρ		RMSE_{CV} (gC/m^2)	
	rainfed	irrigated	rainfed	irrigated	rainfed	irrigated
Cereal	0.41	0.41	0.60	0.64	16.72	14.94
Rice	0.39	0.37	0.62	0.78	14.71	24.02
Maize	0.35	0.45	0.74	0.80	17.79	20.45
Oil[1]	0.51	0.49	0.73	0.81	12.34	7.65
Groundnut	0.24	0.40	0.62	0.79	12.04	17.72

[1] Oil=yield_{max}[soyabean, rapeseed, sunflower].

yield of most crops. Taking management level 4 and the last decade (2085-2094) as an example, for cereal a decline in yield of 6% becomes a growth of 18%; rice and oil show improvements of 37% and 25%, respectively, while groundnut has an increase of 32%. Maize exhibits a weak sensitivity to CO_2, with the globally averaged yield increasing by 11%.

When there is no CO_2 fertilization, with all management levels there is a fairly steady reduction in yield for cereal, rice, maize and groundnut while oil shows an increase in yield. When there is CO_2 fertilization, maize and groundnut yield still change comparatively little over the decades and the effect of random variation is more apparent in their time series plots.

Table 2 summarises the cross-validated performance of the emulators for stage 1 and stage 2 using equation 3.2, for both rainfed and irrigated crops under UKMO-HADGEM1. Maize and oil cross-validated noticeably better than other crops with 74%/73% variance explained when rainfed and 80%/81% explained when irrigated.

Generally, the emulator of the irrigated crops performed better than the emulator of the rainfed crops. This could be attributed to the water stress in rainfed locations, difficult to model, that could complicate the predictions. Stage 1 explained less than 50% of the variance except for rainfed oil. However, the second stage of the algorithm improved results for both the rainfed and irrigated crop systems. For the rainfed crops, the values of variance explained increased from (24-51%) to (60-74%) for all the crops. For the irrigated crops, the first stage explained variance was between 37-49% for all the crops and this increased to 64-81%, as shown in Table 2.

The last two columns of Table 2 show the computed RMSE_{CV} for all scenarios, time slices and management levels, which further examines the accuracy of the emulators. RMSE_{CV} is the difference between the LPJmL and emulator predictions and provides a measure of uncertainty associated with the emulator. Irrigated oil and rice have the lowest and highest values with 7.65 and 24.02 gC/m^2 respectively; a low value indicates more accurate predictions.

The cross validation of stage 2 predictions for four additional GCMs are shown in Table 3. We can see that the emulators again performed well, with results that are typically a little better than in Table 2. The results for CCSR-MIROCMED are better than for other GCMs; this was to be expected because a very similar GCM, CCSR-MIROC32HI, gave part of the training data. Results for irrigated crops are generally better than for rainfed crops, which was also the case in Table 2, though

Table 3: Cross-validated proportion of variance ρ for four GCMs, with CO_2 fertilization, management level 5, RCPs 4.5 and 8.5, and all time slices

Crop	CCSR-MIROCMED		GISS-MODELER		GISS-MODELEH		IPSL-CM4	
	rainfed	irrigated	rainfed	irrigated	rainfed	irrigated	rainfed	irrigated
Cereal	0.79	0.82	0.67	0.62	0.72	0.80	0.74	0.83
Rice	0.79	0.91	0.75	0.77	0.68	0.85	0.75	0.88
Maize	0.86	0.85	0.72	0.66	0.75	0.76	0.83	0.87
Oil	0.88	0.93	0.69	0.78	0.72	0.84	0.81	0.82
Groundnut	0.78	0.89	0.70	0.69	0.70	0.82	0.71	0.84

[1] Oil=yield$_{max}$[soyabean, rapeseed, sunflower].

the results with GISS-MODELER for cereal, maize and groundnut are an exception.

We now consider the spatial comparison between LPJmL and the emulators of UKMO-HADGEM1 for temperate cereal. Figure 4 displays the change in yield between the baseline period (2005-2014) and the decade 2085-2094. The left-hand and right-hand plots show results for LPJmL and the emulator, respectively, with results for rainfed temperate cereal in the upper two plots and those for the irrigated crop in the lower two plots. The emulator under-predicts yield change across the United States for rainfed cereal and over-predicts yield of the irrigated crop in some regions, especially in Eastern Asia and Europe. Overall, the emulator reproduces the global patterns well, especially for the irrigated crop ($\rho = 0.69$ for rainfed cereal and 0.74 for irrigated cereal).

Figure 5 shows results for the rice emulator. The emulator for the rainfed crop under-

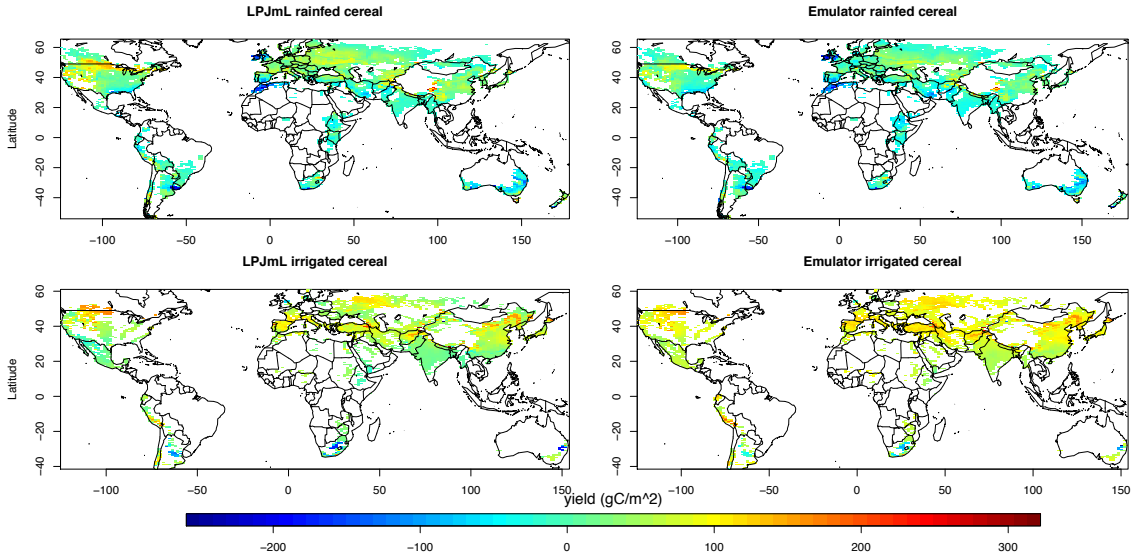

Figure 4: Plots corresponding to mean decadal change in yield for rainfed (top) and irrigated (bottom) cereal for management level 5, RCP6 with CO_2 fertilization and UKMO-HADGEM1. The left-hand plots are the values given by LPJmL and the right-hand plots are the estimates given by our new emulator under cross-validation (c.f. section 4). Areas in white over land correspond to grid cells with observations of zero.

predicts the rice almost everywhere except in Eastern Asia where it reproduces the pattern quite well. On the other hand, the emulator for the irrigated rice reproduces the yield better than the rainfed ($\rho = 0.74, 0.90$ rainfed and irrigated respectively). There is a potential for higher irrigated rice yield in Europe and Asia as shown by both the LPJmL and irrigated emulator plots. Higher yield changes are more prominent with irrigated rice than for rainfed. More irrigated rice is grown than rainfed rice especially in latitude $\geq 30°$, an area that is also characterized with a high change in yield.

Figures for other crops are presented in Supplementary Material S4. Overall, rainfed

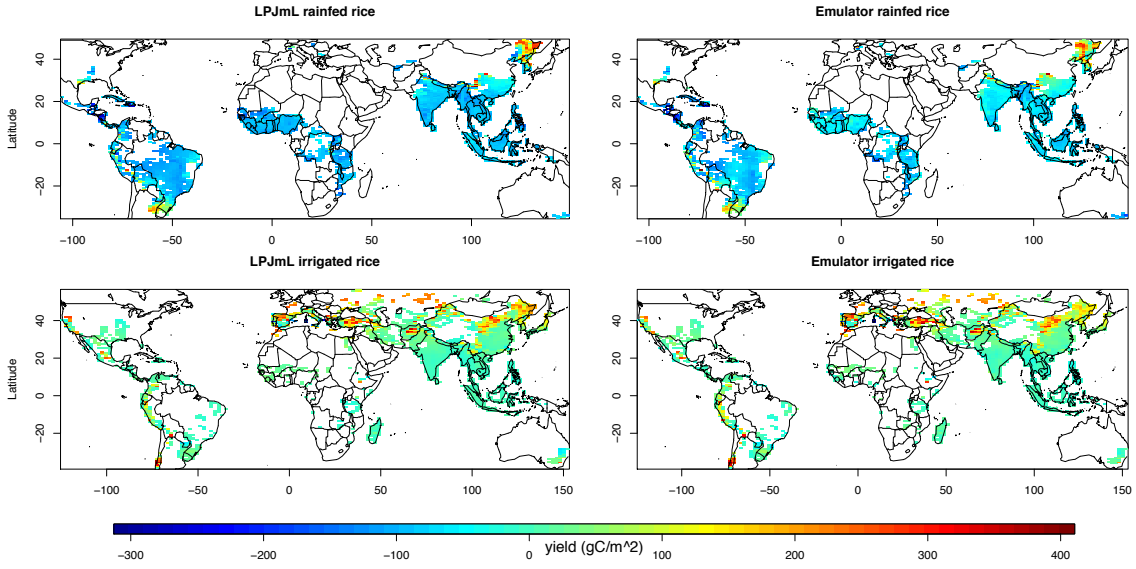

Figure 5: Comparison of LPJmL and the emulator for rainfed and irrigated rice. The plots show the mean decadal change in yield between (2085-2094) and (2005-2014) for management level 5, RCP6, with CO_2 fertilization, for UKMO-HADGEM1.

crop patterns are quite different from the irrigated, as expected, because irrigation allows some crops to be grown where they would not have grown naturally (e.g. rice is grown in Europe with irrigation). Also, more negative changes are prominent in both the LPJmL and emulator predictions for rainfed rice than for other crops. We can clearly see that the emulators cross-validated well as indicated by their ρ values (Tables 2 and 3) and thus captured relatively well the spatial patterns of LPJmL. The maps show visually that the emulator produces patterns that are quite similar to the LPJmL patterns, although in some instances there are over- and under-predictions.

4.2 Sensitivity results

Here, we investigate how the uncertainty in the crop-yield can be partitioned to the various uncertainties in the input variables. We sampled 20 000 observations

directly from the simulation with the CO_2 effect for each of the 37 input variables in Table 1. We computed both the first order (results are not shown) and total sensitivity indices, as described in section 3.3. Bootstrapping was used to compute 95% confidence intervals on the estimated indices. This procedure was applied to all the five crops for both rainfed and irrigated crops. The "sensitivity" package in R (2013) was used for this analysis. Total sensitivity results are shown in Figure 6 for rainfed crops. Results for irrigated crops are given in Supplementary Material S5.

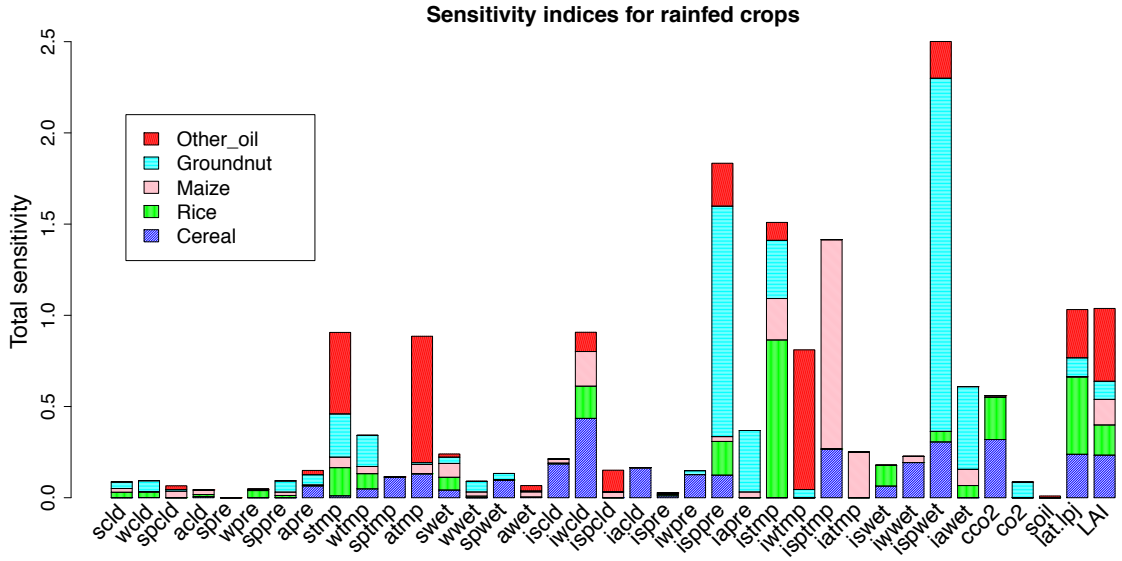

Figure 6: Barplots showing the sensitivity indices for the five rainfed crops over all time-slices, RCPs and GCMs (negative indices are set to 0).

The six most relevant parameters for prediction of temperate cereal yields are initial winter cloud cover, CO_2 change, initial spring wetday frequency, initial spring temperature ('initial' represents baseline value), latitude and LAI, with indices of 0.43, 0.32, 0.30, 0.27, 0.23 and 0.23, respectively. The parameter with the second highest rank is CO_2 change, reflecting the sensitivity of yield to the CO_2 fertilization effect. Atmospheric CO_2 improves crop productivity by stimulating photosynthesis,

thus increasing the number of fruits, seeds etc. The fourth most sensitive parameter is temperature, which generally has a considerable effect on plant growth.

Initial summer temperature, latitude, CO_2 change, LAI and initial spring precipitation are the five most important variables for rice. Their indices are 0.86, 0.42, 0.23, 0.18 and 0.18, making it very clear that summer temperature is the most important parameter for rice. It is twice as influential as latitude (although temperature and latitude are obviously correlated) and four times more influential than CO_2 change.

Rice is a C_3 plant that utilizes direct carbon fixation of CO_2, so CO_2 change is expected to be an important parameter.

The most relevant variables for maize are initial spring, autumn and summer temperature, as well as initial winter cloud cover and LAI. Their indices are 1.1, 0.25, 0.22, 0.19 and 0.14, respectively. It is unsurprising that, as these results show, seasonal temperatures play a key role in the growth and development of maize plants. Maize is less affected by CO_2 fertilization because it is a C_4 plant which has a more efficient mechanism to transport CO_2 to photosynthetic paths. Oil has the following important variables: initial winter temperature, autumn and summer temperature change, LAI and latitude with indices of 0.85, 0.78, 0.40, 0.26 and 0.20, respectively.

Overall, we can see from Figure 6 that baseline seasonal temperature, spring precipitation and wetday frequency, LAI, latitude, CO_2 change and cloud cover are the most important variables across the five crops. Although CO_2 change has a negligible effect on maize, groundnut and oil, our results further support the joint interactive effect of elevated CO_2 and temperature on crop-yield. Some variables are less important (examples are change in seasonal precipitation and cloud cover) and some are unim-

portant (such as summer precipitation and soil) because their calculated sensitivity indices are very low. Baseline CO_2 is of limited importance in this analysis because its level is represented by a single global value and does not vary with time-slice, RCP, or GCM. Similarly, soil is represented in this analysis by discrete values that range from 1-8 across the globe and these are constant across RCP and GCM scenarios.

In general, the sensitivity analysis clearly indicates that temperature is the most important parameter in crop yield projection. This high sensitivity of (growing season) temperature was earlier observed as the major determinant of crop yield change under future climate by Lobell and Burke (2008) and Osborne *et al.* (2013).

5 Concluding comments

This paper has addressed the joint emulation of the impact of climate change, CO_2 fertilization effect and crop management levels on global crop yields. We have described a procedure for constructing an emulator for LPJmL simulations of potential crop yield for cereal, rice, maize, groundnut and oil crop functional types. Two emulators were built for each crop, one for the rainfed crop and the other for its yield under irrigation. Each emulator was constructed using a two-stage process. The first stage used OLS to fit crop-yield as a smooth function of climate variables under the assumption that each spatial point is an independent sample. The second stage involves interpolation of the spatial residuals of unknown scenarios from known scenarios (an approach similar to pattern scaling) using a combination of WLS and PCA. We made similar assumptions to those of GP emulators (O'Hagan, 2006). In particular, it is assumed that the emulator is a smooth and continuous function of its

input variables. The second stage indeed improves the overall predictions.

We used cross-validation to test the accuracy of the emulator and performed sensitivity analysis for each crop response. LPJmL uses daily time-steps as crop-yields respond to daily variability. Here, we have chosen to use only seasonal mean climatic variables as input, as these are readily available input data. With these inputs, under cross-validation the emulators explained 62-93% of the variance for irrigated crops and 60-88% of the variance for rainfed crops. Sensitivity analysis indicated that the predicted yield of rainfed crops depends most heavily on baseline seasonal temperature, CO_2 change, latitude and LAI. Irrigated crops are dominantly sensitive to temperature and less dependent on precipitation, as expected. We provided spatial plots to visually compare the performance of the emulators with the LPJmL simulations.

LPJmL crop data are characterized by a large proportion of zero observations. Grid points where particular crops are not currently grown are represented as zeros in the simulation. A censored regression approach (Moore *et al.*, 2000) was also used to model these data by treating the zero observations as censored observations (results not shown). However, this method was not helpful because of the large dataset we have, coupled with the fact that the censoring algorithm takes a longer computation time, even after reducing the data size from $0.5^\circ \times 0.5^\circ$ to $2^\circ \times 2^\circ$ resolution.

Nevertheless, for the first stage we analysed a small sample of this data with censored regression but the results were not better than OLS. Using either non-linear regression or a Gaussian process emulator would be challenging because of the large number of parameters to be estimated with samples from many scenarios (time-slice, RCP, GCM, management levels, irrigation regime).

We might also have used a dynamic emulation by emulating each grid-cell individually as a function of time. Rather, our choice of OLS is driven by its simplicity.

Our emulation approach extends the work of Lobell and Burke (2010, 2008), which models temporal and spatial variation to predict future crop yields from climate variables and performs sensitivity analyses to examine the importance of temperature and precipitation on future yields. However, unlike Lobell and Burke (2010, 2008), our predictions were not based solely on climate variables but also incorporated soil, latitude, crop management levels and other covariates. Our analyses and aims are also broader; Lobell and Burke (2010) work with just 94 crop-region combinations and only examine temperature and precipitation while Lobell and Burke (2008) work solely with the yield of maize in 200 sites in Sub-Saharan Africa. Our emulators provide estimates of projected change in crop yields at any level of CO_2 emission on high spatial gridded resolutions for five different crops. We also performed rigorous and extensive cross validation for several climate models and RCPs, and our global sensitivity analysis measures the individual contribution of a number of different variables to overall uncertainty. In agreement with Lobell and Burke (2008), the clearest result from our sensitivity results is that temperature is the dominant source of uncertainty in future impacts assessment. The sensitivity analysis used a variance based decomposition technique that handles correlated variables, and the contributions of other variables that it quantifies have not previously been reported.

The emulators reduced the LPJmL model down to a two-stage process that is capable of predicting global crop-yields of different crop functional types, and the spatial distribution and temporal dynamics of these yields in response to a changing climate. These emulators are much faster to run compared to the LPJmL model. LPJmL is

computationally expensive to run, while these emulators give results almost instantaneously. It takes about $8-10$ minutes to simulate eight decadal changes of crop-yield data (on $0.5^\circ \times 0.5^\circ$ resolutions) with a 24G ram, 4-cores Window machine. This approximately 60-fold increase in computational efficiency is particularly useful when the model is coupled to one or more models for the integrated assessment of climate impact.

Acknowledgements

The work was supported by the European Union through FP7 ERMITAGE grant no. 265170.

Appendix 1: Calculations in stage 2

Predictions given by stage 1 for the 16 training scenarios form the $N \times 16$ matrix $\tilde{\mathbf{Y}} = (\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{16})$. Usually a principal component (PC) analysis would be applied to $\tilde{\mathbf{Y}}$: in the standard data analysis problem, each column of $\tilde{\mathbf{Y}}$ would relate to a different variable and the aim would be to condense the different variables into a few components. Here, each column of $\tilde{\mathbf{Y}}$ corresponds to a different scenario and these different scenarios must retain their identities. However, we want to condense the information given by the N grid points into a few summary statistics. Thus, we want to reduce the number of rows (rather than the number of columns) so a principal component analysis is applied to $\tilde{\mathbf{Y}}^{\mathrm{T}}$. The spectral decomposition theorem gives

$$\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{\mathrm{T}} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^{\mathrm{T}} \tag{A.1}$$

where $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_N)$ is a diagonal matrix of eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$, and $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_N)$ is an $N \times N$ orthogonal matrix, where $\boldsymbol{\gamma}_k$ is the eigenvector corresponding to λ_k. The rank of $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{\mathrm{T}}$ is 16 (or possibly less), so λ_{17}, λ_{18} ... and λ_N each equal 0.

The PC transformation from $\tilde{\mathbf{y}}$ to $\tilde{\mathbf{x}}$ is given by

$$\tilde{\mathbf{y}} \rightarrow \mathbf{\Gamma}^{\mathrm{T}}\tilde{\mathbf{y}} = \tilde{\mathbf{x}}. \tag{A.2}$$

Put $\tilde{\mathbf{x}}_j = \mathbf{\Gamma}^{\mathrm{T}}\tilde{\mathbf{y}}_j$ for $j = 1, \ldots, 16$ and put $\tilde{\mathbf{x}}^\star = \mathbf{\Gamma}^{\mathrm{T}}\tilde{\mathbf{y}}^\star$, where $\tilde{\mathbf{y}}^\star$ are the stage 1 predictions for the new scenario. Then $(\tilde{\mathbf{y}}^\star - \tilde{\mathbf{y}}_j)^{\mathrm{T}}(\tilde{\mathbf{y}}^\star - \tilde{\mathbf{y}}_j) = (\tilde{\mathbf{x}}^\star - \tilde{\mathbf{x}}_j)^{\mathrm{T}}(\tilde{\mathbf{x}}^\star - \tilde{\mathbf{x}}_j)$. We wish to give more importance to the eigenvectors that corresponds to large eigenvalues, so we define the distance between $\tilde{\mathbf{y}}^\star$ and $\tilde{\mathbf{y}}_j$ as

$$d_j = \left[\sum_{k=1}^{N} \lambda_k (x_k^\star - x_{jk})^2\right]^{1/2} = \left[\sum_{k=1}^{16} \lambda_k (x_k^\star - x_{jk})^2\right]^{1/2}, \tag{A.3}$$

where $x_k^\star$ and x_{jk} are the k^{th} components of $\tilde{\mathbf{x}}^\star$ and $\tilde{\mathbf{x}}_j$, respectively.

In forming regression equations, we use only the first four principal components; for every crop/irrigation regime/management level/time slice combination, these explained at least 95% of the variation in $\tilde{\mathbf{Y}}$.

Let $\hat{\mathbf{\Gamma}}$ be the first four columns of $\mathbf{\Gamma}$, so $\hat{\mathbf{\Gamma}} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3, \boldsymbol{\gamma}_4)$. We put $\mathbf{X}_0 = \tilde{\mathbf{Y}}^{\mathrm{T}}\hat{\mathbf{\Gamma}}$.

Let $\mathbf{y}_j$ be the $N \times 1$ vector of observation given by LPJmL for the j^{th} scenario of the training set ($j = 1, \ldots, 16$); $\tilde{\mathbf{y}}_j$ is the corresponding vector of prediction given by stage 1, and $\boldsymbol{\varepsilon}_j = \mathbf{y}_j - \tilde{\mathbf{y}}_j$ gives the error in the stage 1 predictions. A weighted least squares (WLS) regression is used to estimate $\boldsymbol{\varepsilon}^\star$, the corresponding error in $\tilde{\mathbf{y}}^\star$. A separate regression equation is determined for each component of $\boldsymbol{\varepsilon}^\star$. We put $\mathbf{E} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{16})$ so for the i^{th} regression (the i^{th} grid cell) the values of the dependent variable are $\tilde{\boldsymbol{\varepsilon}}_i^{\mathrm{T}}$,

where $\tilde{\boldsymbol{\varepsilon}}_i^{\mathrm{T}}$ is the i^{th} row of $\mathbf{E}$.

If any distance d_j is 0, then weighted regression is unnecessary as the scenario with the zero-distance gets a weight of infinity, and the errors for that scenario are taken as the errors in $\tilde{\mathbf{y}}^{\star}$. Sometimes the distance equals 0 for more than one scenario and then the error for those scenarios are averaged. Specifically, if Q denotes the set of integers such that $d_j = 0$ for $j \in Q$, then the vector of estimate error for the new scenario is $\hat{\boldsymbol{\varepsilon}}^{\star} = \sum_{j \in Q} \boldsymbol{\varepsilon}_j$.

For non-zero d_j, the weights w_j $(j = 1, \ldots, 16)$ and weight matrix $\mathbf{W}$ are defined as

$$w_j = \frac{(1/d_j^2)}{\sum_1^{16}(1/d_j^2)} \qquad \text{and} \qquad \mathbf{W} = \begin{pmatrix} w_1 & 0 & \ldots & 0 \\ 0 & w_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & w_{16} \end{pmatrix}.$$

We take one grid point at a time and form a separate regression equation for that grid point. The data for one of these regressions is the (16×1) vector of responses $\tilde{\boldsymbol{\varepsilon}}_i$ (the errors at that grid point) and the (16×4) matrix $\mathbf{X}_0$, which holds the values taken by the explanatory variables.

The weighted linear regression uses $\mathbf{W}$ as the weight-matrix.

For the i^{th} grid point, the regression model is $E(\boldsymbol{\varepsilon}_i \,|\, \mathbf{x}_0) = \boldsymbol{\beta}_i^{\mathrm{T}} \mathbf{x}_0$ where $\boldsymbol{\varepsilon}_i$ is a random variable whose observed value is the j^{th} component of $\tilde{\boldsymbol{\varepsilon}}_i$ when $\mathbf{x}_0^{\mathrm{T}}$ is the j^{th} row of $\mathbf{X}_0$. The WLS estimate of $\boldsymbol{\beta}_i$ is

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_0^{\mathrm{T}} \mathbf{W} \mathbf{X}_0)^{-1} \mathbf{X}_0^{\mathrm{T}} \mathbf{W} \tilde{\boldsymbol{\varepsilon}}_i. \tag{A.4}$$

For the new scenario, put $\mathbf{x}_0^{\star} = \hat{\boldsymbol{\Gamma}}^{\mathrm{T}} \tilde{\mathbf{y}}^{\star}$. Then the estimate of the prediction error in $\tilde{\mathbf{y}}^{\star}$ at the i^{th} grid point is

$\varepsilon_i^\star = \hat{\boldsymbol{\beta}}_i^{\mathrm{T}} \mathbf{x}_0^\star,$

which (using (A.4)) can be written as

$$\hat{\varepsilon}_i^\star = \tilde{\boldsymbol{\varepsilon}}_i^{\mathrm{T}} \mathbf{W}\mathbf{X}_0 (\mathbf{X}_0^{\mathrm{T}} \mathbf{W}\mathbf{X}_0)^{-1} \mathbf{x}_0^\star. \tag{A.5}$$

Equation (A.5) estimates the prediction error separately for each grid cell. To combine the calculations for all grid cells into a single step, let $\hat{\boldsymbol{\varepsilon}}^\star = (\hat{\varepsilon}_1^\star, \ldots, \hat{\varepsilon}_N^\star)^{\mathrm{T}}$. Then the equation for estimating all the residuals in $\tilde{\mathbf{y}}^\star$ is given as

$$\hat{\boldsymbol{\varepsilon}}^\star = \mathbf{E}\mathbf{W}\mathbf{X}_0 (\mathbf{X}_0^{\mathrm{T}} \mathbf{W}\mathbf{X}_0)^{-1} \mathbf{x}_0^\star. \tag{A.6}$$

The similarity of our method to pattern scaling stems from this equation. The vector $\hat{\boldsymbol{\varepsilon}}^\star$ is the estimated error pattern for the new scenario and $\boldsymbol{\varepsilon}_j$ is the (known) error pattern for the j^{th} training scenario. If we put $\mathbf{W}\mathbf{X}_0 (\mathbf{X}_0^{\mathrm{T}} \mathbf{W}\mathbf{X}_0)^{-1} \mathbf{x}_0^\star = (\alpha_1, \ldots, \alpha_{16})^{\mathrm{T}}$, then (A.6) may be written as

$$\hat{\boldsymbol{\varepsilon}}^\star = \sum_{j=1}^{16} \alpha_j \boldsymbol{\varepsilon}_j,$$

so the estimated error pattern for the new scenario is simply a linear combination of the error patterns of the training scenarios.

To avoid extrapolation, we bound the i^{th} component of $\hat{\boldsymbol{\varepsilon}}^\star$ to be within the range of $\tilde{\boldsymbol{\varepsilon}}_i$. Let $\hat{\boldsymbol{\varepsilon}}^\#$ denote the resulting vector when a component of $\hat{\boldsymbol{\varepsilon}}^\star$ is set equal to any bound it exceeds, as detailed in equation (3.1). We take $\hat{\mathbf{y}}^\star = \tilde{\mathbf{y}}^\star + \hat{\boldsymbol{\varepsilon}}^\#$ as the emulated value of $\mathbf{y}$ for the new scenario.

Appendix 2: Variograms and distance metrics

Distance weighted regression is used to estimate the residual pattern of an unknown scenario from the scenarios with known residual patterns (Section 3.1.2). More weight

is assigned to known scenarios that are similar in pattern to the unknown scenario. The distance, d_j, between the unknown jth (known) scenario is defined in equation (A.3) and is taken as the measure of pattern similarity.

Three simple metrics for converting the d_j distances to weights were considered: linear ($w_j \propto d_j^{-1}$), quadratic ($w_j \propto d_j^{-2}$), and cubic ($w_j \propto d_j^{-3}$), where the weights (w_j) are scaled so that $\sum w_j = 1$. The use of covariance functions to determine weights was also explored using variograms. For each crop/irrigation regime/time slice/ management level there are sixteen different training scenarios that predict the yield in a grid cell. Generalising equation (A.3), let $d_{jk} = \left[\sum_{m=1}^{16} \lambda_m (x_{jm} - x_{km})^2\right]^{1/2}$ denote the distance between the jth and kth scenarios, and let z_{ijk} denote the difference between their predictions in grid cell i.

To obtain the empirical variogram, we split the range of d_{jk} into sections (bins) of equal length. The empirical variogram, $\hat{\eta}(d)$ is defined as (Cressie, 1993)

$$\hat{\eta}(\overline{d}_\ell) = \frac{1}{M_\ell N} \sum_{(j,k)\in\Omega_\ell} \sum_{i=1}^{N} |z_{ijk}|$$

where $\overline{d}_\ell$ is the middle of bin ℓ, $(j,k) \in \Omega_\ell$ if d_{jk} is in bin ℓ, and M_ℓ is the number of items in Ω_ℓ. In Figure 7 the small circles are the values of $\hat{\eta}(\overline{d}_\ell)$ at the midpoint of each bin. The lines in the figure show the theoretical variograms for the following covariance models.

Spherical covariance model: $$\eta(d) = \begin{cases} \sigma^2 \left(\frac{3d}{2\phi} - \frac{d^3}{2\phi^3}\right) & \text{if } d \le \phi \\ \sigma^2 & \text{if } d > \phi \end{cases}$$

Power exponential covariance: $\eta(d) = \sigma^2\{1 - \exp(-3d/\phi)^\kappa\}, 0 < \kappa \le 2.$

Gaussian covariance model: $\eta(d) = \sigma^2\{1 - \exp(-3d^2/\phi^2)\}$

Matern covariance model: $$\eta(d) = \sigma^2 \left\{1 - \frac{(\phi d)^\kappa \mathcal{B}_\kappa(\phi d)}{2^{\kappa-1}\Gamma(\kappa)}\right\}.$$

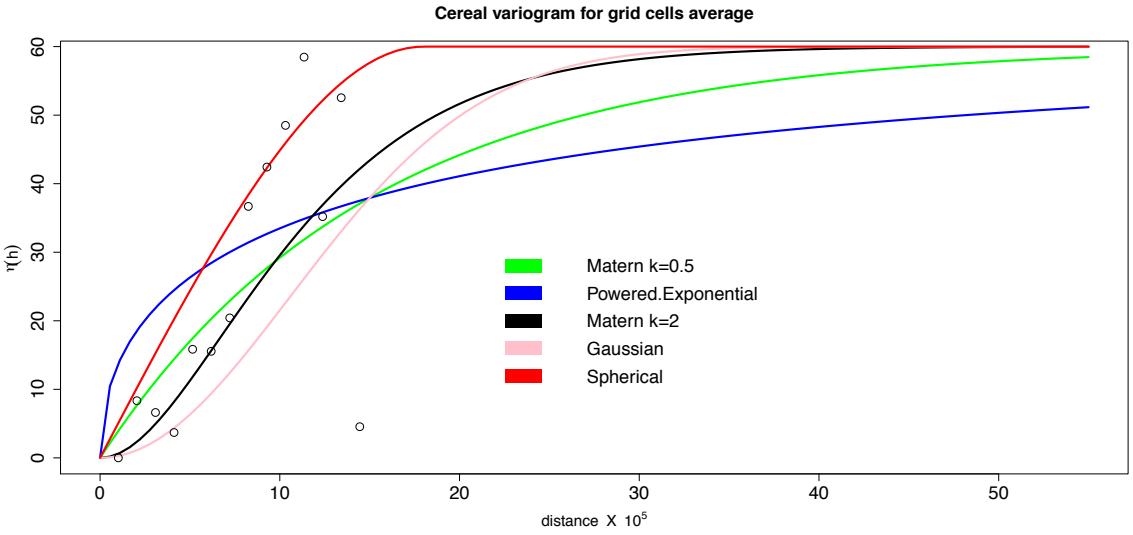

Figure 7: Empirical and theoretical variogram for rainfed cereal. The points are the estimated variogram bins using the residual data, while the curves are the theoretical models fitted using various covariance models.

The parameters σ and ϕ are estimated from the empirical variogram. For the Matern model, κ must be specified and we consider two values, $\kappa = 0.5$ and $\kappa = 2$; $\mathcal{B}_\kappa$ is the modified Bessel function of the second kind of order κ.

The plotted points in Figure 7 show that $\hat{\eta}(\overline{d}_\ell)$ increases rapidly as $\overline{d}_\ell$ increases from 0 (and then levels off), so it is clear that errors for the different scenarios are correlated and the correlation increases as distance reduces. Variogram plots for different crops and some randomly chosen grid cells are given in Supplementary Material S6. These show the same trend. For all crops the theoretical covariance models follow the empirical points reasonably closely with no model being clearly the best.

A covariance model assigns weights to scenarios so that $w_j \propto \{\eta(d_j)\}^{-1}$ where, as with the other metrics, weights are scaled so that $\sum w_j = 1$. The emulator was fitted using different weight functions, enabling the covariance models and other metrics to

Table 4: Cross-validated proportion of variance ρ for various covariance functions used as a weight in WLS fitting compared to quadratic metric with management level 5, RCP 6 between (2085-2094) and (2005-2014).

Function	Cereal		Rice		Maize		Oil[1]		Groundnut	
	rainfed	irrig[2]	rainfed	irrig	rainfed	irrig	rain	irrig	rainfed	irrig
Quadratic	0.69	0.74	0.74	0.90	0.81	0.86	0.80	0.88	0.68	0.79
Matern $\kappa = 0.5$	0.78	0.82	0.78	0.84	0.84	0.87	0.79	0.88	0.67	0.79
Spherical	0.76	0.82	0.78	0.84	0.83	0.73	0.75	0.84	0.68	0.79
Matern $\kappa = 2$	0.78	0.82	0.80	0.85	0.85	0.88	0.79	0.88	0.68	0.81
Matern $\kappa = 0.1$	0.79	0.82	0.78	0.83	0.83	0.88	0.79	0.88	0.67	0.79
Power expo[3]	0.76	0.81	0.65	0.89	0.79	0.86	0.45	0.82	0.59	0.85
Gaussian	0.76	0.80	0.76	0.68	0.77	0.56	0.70	0.79	0.66	0.75

[1] Oil=yield$_{max}$[soyabean, rapeseed, sunflower], [2] "irrig" denotes irrigated crop.

[3] "Power expo" denotes Power exponential.

be further compared using cross-validation. Results are presented in Table 4. The power exponential covariance model led to poor predictions for rainfed oil and rainfed groundnut, but otherwise all the methods of choosing weights led to reasonably good predictions. The quadratic distance metric is much simpler than the other metrics in Table 4 and it explained 68-90% of the variation in the LPJmL predictions, compared with 45-89% for the covariance models. Because of its simplicity and comparable performance, we chose it as the means of determining weights for the WLS regression in stage 2 of the emulator algorithm.

References

Arnell NW, Lowe JA, Brown S, Gosling SN, Gottschalk P, Hinkel J, et al. (2013) A global assessment of the effects of climate policy on the impacts of climate change. Nature *Climate Change*, **3**, 512–519.

Bondeau A, Smith PC, Zaehele S and Schaphoff S et al. (2007) Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biology*, **13**, 1–28.

Bornn L and Zidek JV (2012) Efficient stabilization of crop yield prediction in the Canadian Prairies. *Agricultural and Forest Meteorology*, **152**, 223–232.

Chiles JP and Delfiner P (1999) *Geostatististics, Modelling Spatial Uncertainty*. New York: John Wiley.

Cressie N (1993) *Statistics for Spatial Data*. New York: John Wiley.

Deryng D, Sacks WJ, Barford CC and Ramankutty N (2011) Simulating the effects

of climate and agricultural management practices on global crop yield. *Global biogeochemical cycles*, **25**.

Ecker MD and Gelfand AE (1997) Bayesian variogram modeling for an isotropic spatial process. *Journal of the Agricultural, Biological and Environmental Statistics*, **2**, 347–369.

Fader M, Rost S, Muller C, Bondeau A. and Gerten D (2010) Virtual water content of temperate cereals and maize: Present and potential future patterns. *Journal of Hydrology*, **384**, 218–231.

Fischer G, Shah M, van Velthuizen H and Nachtergaele FO (2001) *Global agroecological assessment for agriculture in the 21st century.* IIASA Research Report 02-02, International Institute for Applied Systems Analysis, Laxenburg, Austria, p. 119.

Handcock MS and Stein ML (1993) Bayesian analysis of kriging. *Technometrics*, **35**, 403–410.

Higdon D, Gattiker J, Williams B and Rightley M (2008) Computer model calibration using high-dimensional input. *Journal of the American Statistical Association*, **103**, 570–583.

Holden PB, Edwards NR, Oliver KIC, Lenton TM and Wilkinson RD (2010) A probabilistic calibration of climate sensitivity and terrestrial carbon change in GENIE-1, Climate Dynamics, **35**, 785–806.

Holden, P.B., Edwards, N.R., Garthwaite, P.H., Fraedrich, K., Lunkeit, F., Kirk, E., ... & Babonneau, F. (2014). PLASIM-ENTSem v1. 0: a spatio-temporal emulator

of future climate change for impacts assessment. *Geoscientific Model Development*, $7(1), 433 - 451$.

IPCC Fifth Assessment Synthesis Report (2014) Edited by R K Pachauri, L Meyer and the Core Writing Team. (Available from `http://www.ipcc.ch/report/ar5/syr`.)

Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt LA. et al. (2003) The DSSAT cropping system model. *European Jounal of Agronomy*. **18**, 235–265.

Kart RW (1979) Sensitivity analysis of statistical crop weather models. *Agricultural Meteorology*, **20**, 291–300.

Kim Y, Lee J and Kim J (2005) Bayesian bootstrap analysis of doubly censored data using Gibbs sampler. *Statistica Sinica*, 969–980.

Leemans R and Solomon A (1993) Modeling the potential change in yield and distribution of the earth's crops under a warmed climate. *Climate Research*, **3**, 79–96.

Lobell DB (2013) Errors in climate datasets and their effects on statistical crop models. *Agricultural and Forest Meteorology*, **170**, 58–66.

Lobell DB and Burke MB (2008) Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation. *Environmental Research Letter*, **3**, 034007.

Lobell DB and Burke MB (2010) On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, **150**(11), 1443–1452.

Lobell DB, Cahill KN and Field CB (2006) Weather-based yield forecasts developed for 12 California crops. *California Agriculture*, **60**(4), 210–214.

Matis JH, Birkett T and Boudreaux D (1989) An application of the Markov chain approach to forecasting cotton yields from Surveys. *Agricultural Systems*, **29**, 357–370.

Meinshausen M, Raper SC and Wigley TM (2011) Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 Part 1: Model description and calibration. *Atmospheric, Chemistry and Physics*, **11**, 1417–1456.

Moore MR, Gollehon NR and Hellerstein DM (2000) Estimating producer's surplus with the censored regression model: An application to producers affected by Columbia River Basin Salmon Recovery. *Journal of Agricultural and Resource Economics*, **25**(2), 325–346.

Moss RH, Edmonds JA, Hibbard KA, Manning MR and Rose SK, et al. (2010) The next generation of scenarios for climate change research and assessment. *Nature*, **463**, 747–756.

Muller C and Robertson R (2014) Projecting future crop productivity for global economic modeling. *Agricultural Economics*, **45**(1), 37–50.

Nelson GC, Valin H, Sands RD, Havlik P, Ahammad, H, Deryng D. et al. (2014) Climate change effects on agriculture: Economic responses to biophysical shocks. *Proceedings of the National Academy of Sciences*, **111**(9), 3274–3279.

O'Hagan A (2006) Bayesian Analysis of Computer Code Outputs: A Tutorial. *Reliability Engineering and System Safety*, **91**, 1290–1300.

Osborn TJ (2009) *A user guide for ClimGen: a flexible tool for generating monthly climate data sets and scenarios*. Climatic Research Unit, University of East Anglia, Norwich, 17.

Osborne T, Rose G and Wheeler T (2013) Variation in global-scale impacts of climate change on crop productivity due to climate model uncertainty and adaptation. *Agriculutural and Forest Meteorology*, **170**, 183–194.

Osborne TM and Wheeler TR (2013) Evidence for a climate signal in trends of global crop yield variability over the past 50 years. *Environmental Research Letter*, **8**, 024001.

Parry ML, Rosenzweig C, Iglesias A, Livermore M and Fischer G (2004) Effects of climate change on global food production under SRES emissions and socio-economic scenarios. *Global Environmental Change*, **14**, 53–67.

R: A Language and Environment for Statistical Computing, R Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2013, ISBN 3-900051-07-0, http://www.R-project.org/.

Reddy VR and Pachepsky YA (2000) Predicting crop yields under climate change conditions from monthly GCM weather projections. *Environmental Modelling and Software*, **15**, 79–86.

Rosenzweig C, Elliott J, Deryng D, Ruane AC, Muller C, Arneth A, et al. (2014) Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences*, **111**(9), 3268–3273.

Saltelli A (2002) Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communication*, **145**, 280–297.

Schabenberger O and Gotway CA (2005) *Statistical Methods for Spatial Data Analysis*. Boca Raton: Chapman & Hall.

Schlenker W and Roberts MJ (2006) *Estimating the impact of climate change on crop yields: The importance of non-linear temperature effects (with discussion)*, 0607-01, Department of Economics Columbia University New York, NY.

Sitch S, Smith B, Prentice IC, Arneth A and Bondeau A, et al. (2003) Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, **9**, 161–185.

Smith P and Gregory PJ (2013) Climate change and sustainable food production. *Proceedings of the Nutrition Society*, **72**(1), 21–28.

Solomon S (2007) (ed) *Climate Change: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovermental Panel on Climate Change. Cambridge: Cambridge University Press.

UNFPA. (2010). *State of world population 2010: From conflict and crisis to renewal: generations of change*. United Nations Population Fund, available from: http://www.unfpa.org/swp/2010/web/en/pdf/EN SOWP10.pdf

Wallach D (2011) Crop model calibration: A statistical perspective. *Agronomy Journal*, **103** (4), 1144–1151.

Warren R, Lowe JA, Arnell NW, Hope C, Berry P, Brown S, et al. (2013) The AVOID programmes new simulations of the global benefits of stringent climate change mitigation. *Climatic Change*, **120** (1-2), 55–70.