

# Bayesian emulation of climate change impact on crop yields

Oluwole K. Oyebamiji · Neil R. Edwards · Philip B. Holden ·  
Paul H. Garthwaite · Sibyll Schaphoff · Dieter Gerten

Received: date / Accepted: date

**Abstract** The global climate is changing and this has a great negative impact on ecosystems. It adversely affects the environment, life quality, the economy, agriculture. Assessment of future climate change on agricultural productivity is attracting increasing attention among the modellers, since modelling provides a thorough way of quantifying the crop yield response under climate uncertainty. Adequate knowledge of the impacts of climate change on agricultural crop is required to fully understand global food security. In order to address this problem, the common tradition is to run a high complex process-based simulation models. This assessment is challenging because the models are computationally intensive and simulation runs are often limited. This problem can be eradicated by using a statistical approximation of the complex models which will help in reducing the computational burden. Our aim in this paper is to build a cheaper surrogate models (called emulators) from simulations of the LPJmL dynamic global vegetation and crop model. This paper has demonstrated emulation technique as a means for assessing the impact of climate change on agricultural productivity. The paper focuses on global projections of crop yield change with their associated uncertainty levels. We proposed a two-stage approach where we used a stepwise regression in the first stage and combined both principal component analysis and Bayesian technique for residual interpolation in the second stage .

**Keywords** Bayesian techniques · stepwise regression · PCA · LPJmLem · crop yields · GP emulator

## 1 Introduction

The global climate is changing and this has a great negative impact on ecosystems. It adversely affects the environment, life quality, the economy and agriculture. Any change in the climate system over time, whether due to natural variability or human activity, is referred to as climate change (IPCC, 2007). Of a particular priority in this paper is the assessment of the impacts of future climate change on terrestrial ecosystems. Assessment of climatic impacts on global vegetation is attracting increasing attention among modellers, since

---

The Open University, Environment, Earth & Ecosystems  
Walton Hall, Milton Keynes,  
MK7 6AA, UK.  
Tel.: (+44) 7411875750  
Fax: (+44) 1908 655 151  
E-mail: wolemi2@yahoo.com

Neil R. Edwards · Philip B. Holden  
The Open University, Environment, Earth & Ecosystems, Milton Keynes, UK ·  
P.H. Garthwaite  
Department of Mathematics & Statistics, The Open University, Milton Keynes, UK ·  
S. Schaphoff · D. Gerten  
Potsdam Institute for Climate Impact Research (PIK), Telegrafenberg A62, 14473 Potsdam, Germany

modelling provides a thorough way of quantifying the biospheric response under climate uncertainty. Adequate knowledge of the impacts of climate change on agricultural crops is required to fully understand global food security. Future increases in the demand for food and reduced availability of land for growing pose major challenges for human society as a result of rising projections for world population over the next century. Growth in demand for food will also put pressure on water resources. Rising temperature, CO<sub>2</sub> emissions and associated climate change will affect global food and water supplies. Potential climatic impacts on vulnerable populations through lack of adequate food and regular water supply have been identified as another serious threats. [IPCC \(2014\)](#) reported that for the major crops (wheat, rice, and maize) in tropical and temperate regions, climate change without adaptation will negatively impact production for local temperature increases of 2°C or more above late-20th-century levels, although individual locations may benefit with medium confidence.

Modelling tools for studying the interaction between climate change and its effects are relatively under-developed. In spite of tremendous progress in the statistical analysis of climate change studies, quantifying the climatic impact on agricultural crops still suffer problems. In addition, adequate knowledge of climatic impacts on agricultural productivity is essential. For instance, comprehensive and detailed understanding of the impact of climate change on global crop yields is required for maintaining good management practices for sustainable agriculture. However, this assessment is challenging and computationally intensive models have to be run.

The Lund-Potsdam-Jena-managed Land (LPJmL) model is a process-based dynamic global vegetation and crop model. The global responses of carbon and vegetation patterns under climate change for both natural and agricultural ecosystems can be simulated by LPJmL. The model is used in this study to simulate the response of terrestrial ecosystems to climate change. The run times for LPJmL are high and this often makes the full assessment of the impact of climate change on the ecosystems impossible. Both statistical and process-based models have been used extensively since fourth assessment results AR4 to assess the response of crop yield to temperature ([IPCC, 2014](#)). Statistical emulation can help to overcome this problem. Emulation integrates both the process-based model and statistical techniques. Besides, emulators offer rapid and relatively quick alternatives for projection of climate change impacts on agricultural productivity for diverse climate scenarios ([Oyebamiji et al., 2015](#)). For instance, [Warren et al. \(2013\)](#), [Arnell et al. \(2013\)](#) considered both economic and physically based modelling approaches to project global and regional impact of climate change, but due to computing constraints only a limited set of scenarios were investigated. Another benefit of emulation is the provision of a measure of uncertainties associated with the projections.

Several techniques have been used for modelling effect of climate and weather on crop yield. [Kitchen et al. \(1999\)](#) compared the predictive performance of stepwise multiple linear regression with that of projection pursuit regression and neural networks. In related studies [Schlenker et al. \(2006\)](#) and [Sacks et al. \(2010\)](#) considered multiple linear regression techniques using polynomial and interaction terms for modelling crop productivity measure. In contrast to the regression approaches of [Wallach \(2011\)](#), [Schlenker et al. \(2006\)](#), [Osborne et al. \(2013\)](#), [Matis et al. \(1989\)](#) and [Jain & Agrawal \(1992\)](#) several researches applied a non-parametric Markov chain approach to crop yield prediction. [Kim et al. \(2005\)](#) applied a Bayesian bootstrap method to similar data.

[Tebaldi et al. \(2005\)](#), [Lopez et al. \(2006\)](#) proposed a Bayesian statistical model that combines information from a multi-model ensemble of Atmospheric Ocean General Circulation Models (AOGCMs) with observations in order to determine the probability distribution of future climate change. A probabilistic method to overcome the challenge of inclusion of model inadequacy in order to improve the quality of a simulator ensemble was presented in [Kennedy et al. \(2008\)](#) and in [Rougier \(2007\)](#). [Gosling et al. \(2007\)](#) estimated the uncertainty associated with the carbon flux simulated from a dynamic global vegetation model. They used a GP emulation technique proposed by [Oakley & O'Hagan \(2002\)](#) to quantify these uncertainties for England and Wales.

Similarly, [Higdon et al. \(2008\)](#) applied the [Oakley & O'Hagan \(2002\)](#) approach in conjunction with a PCA for basis representations of high-dimensional output. Apart from reducing the dimensionality of the problem

this PCA technique also reduces the computation time required for obtaining the posterior distributions. A closely related study of [Wilkinson \(2010\)](#) performed a calibration of multivariate experiments by extending the approach of [Kennedy et al. \(2001\)](#) to multivariate models. This study incorporated principal component analysis, like [Higdon et al. \(2008\)](#), to project the multivariate output to a lower subspace. [Boukouvalas et al. \(2008\)](#) focuses on variety of dimension reduction methods. They illustrated the methodologies by modelling time series output data.

[Young et al. \(2011\)](#) described the behaviour of large linear dynamic models using statistical principles of dynamic emulation. Their approach identifies a low-order dynamic model that approximates the behaviour of the high-order dynamic simulator at a low computational cost. [Kennedy et al. \(2001\)](#) applied a Bayesian technique to calibrate computer models. They obtained the posterior distribution of a parameter that measured residual uncertainty. [Oakley & O'Hagan \(2004\)](#) described a Bayesian method for quantification of uncertainty in complex computer models. [Kennedy et al. \(2006\)](#) described some notable examples where GP modelling applications have been implemented. [Heitmann et al. \(2006\)](#) and [Higdon et al. \(2008\)](#) applied a principal component to decompose and emulate multivariate responses. The goal of this paper is to assess the global impact of climate change on agricultural crops and to provide measure of uncertainty associated with the projections.

We proposed a two-stage approach where we used a stepwise regression in the first stage and combined both principal component analysis and Bayesian technique for residual interpolation in the second stage. We first applied a linear regression to the multi-dimensional and high spatial resolution data. This will capture most of the spatial and temporal variability. The residuals obtained from this regression results are aggregated to a country level and modelled by a combination of PCA and GP regression. In the second stage, we used a Bayesian technique for emulation of unexplained residual. The final results are then compared with [Oyebamiji et al. \(2015\)](#) approach that demonstrated a weighted least regression for residual interpolation.

## 2 Materials and methods

### 2.1 Models used for the analysis

The data used for this study are provided by the LPJmL and Spatial Climate Generator (ClimGEN).

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. 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  $0.5^\circ$  resolution grid cells. ClimGEN emulates the spatial response patterns by disaggregating the temperature trajectories to  $0.5^\circ$  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.

### 2.2 Simulation data

The input data are monthly climate variables (temperature, precipitation, cloud cover and wetday frequency) from ClimGEN. All the simulation data are on a global  $0.5^\circ \times 0.5^\circ$  degrees resolution. The data cover a period of 2001 – 2100. There are seven different General Circulation Models (GCMs) and each is applied to 4 RCPs, giving a total of 28 different climate scenarios. The other inputs are annual  $\text{CO}_2$  concentrations for all the 4 RCPs, soil data that describes 8 different soil types.

The output data from LPJmL are potential crop yields. The crop yields are annual values for 12 CFTs, with rainfed and irrigated outputs considered separately. We have chosen only 4 CFTs namely temperate

cereal, rice, maize and oil, (oil crop is a combination of sunflower, soyabean and rapeseed) for emulation. In addition, the crop yields have 7 seven different crop management levels for each scenario, and performed simulations with and without CO<sub>2</sub> fertilization effect. 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  $LAI_{max}$ .

### 3 Methods

A Bayesian framework for emulation is almost always based on the assumption that a Gaussian process prior distribution can be specified for unknown parameters and hyperparameters. Under a Bayesian perspective unknown parameters are treated as random variables. The given prior distribution can be updated from training data. Applying Bayes rule to this setting, a posterior distribution can be obtained. The posterior distribution is also a Gaussian process. A major difficulty with GP modelling is the computational effort associated with dealing with large data, as computer time scales are of order  $O(n^3)$  where  $n$  is the number of observations. Several techniques have been adopted to overcome this computational problem. Earlier techniques are documented in [Rasmussen & Williams \(2006\)](#) and [Quinonero-Candela & Rasmussen \(2005\)](#). In particular, [Shi et al. \(2005\)](#) used hierarchical Gaussian process mixtures for regression of a large dataset with repeated measurements. GP emulation is based on the Bayesian technique and experimental design of computer experiments for predicting model outputs at test input point [Sacks et al. \(1989\)](#) and [\(Santner et al., 2003\)](#). A GP emulator assumes that a simulator output is an unknown function  $g(\cdot)$ . We can then choose a prior distribution for  $g(\cdot)$  using the Bayesian approach and update this distribution, with some data obtained from the simulator runs.

There are several techniques for estimating hyperparameters of the GP prior. The maximum likelihood, restricted maximum likelihood and MCMC methods are the standard ones. Maximum likelihood is computationally feasible and depends on the assumption of multivariate Gaussian distribution. Analytical estimates of the hyperparameters of  $\beta$  and  $\sigma_s^2$  by integrating out the nuisance parameters are discussed in the next section. We then apply the MLE technique for the estimation of correlation  $\rho$  hyperparameter as described in Appendix 1.

#### 3.1 Gaussian process emulation

In this paper, we analyse the crop yield data using Bayesian techniques. We are implementing a Bayesian technique because of its wide applicability and flexibility. Another additional benefit of Bayesian modelling is the quantification of the model uncertainty. The motivation for this paper is to demonstrate the concept of GP regression in modelling crop yield data and to compare the results with some existing technique. In particular, we demonstrate combinations of stepwise regression, PCA and GP techniques in emulating the crop yield response under global climate scenarios. We examine whether the non-parametric modelling of the unexplained residual using Gaussian process (GP) in the second stage provides a useful result. LPJmL simulation runs are computationally expensive and to reduce this high computational cost we described the statistical emulation of LPJmL crop outputs in this paper, where a two-stage technique was proposed. In the first stage of that paper, stepwise regression was applied onto crop yield, and a GP regression was used for residual interpolation in the second stage.

A Gaussian process model could not be applied directly to the first stage because of the computational difficulty in the 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 decadal and sampling from each scenario. The data matrix contains over 4.5 million observations. It is possible to use GP for residual interpolation. However, this approach still has a high computational cost, and it is necessary to reduce the spatial resolution and aggregate data to a country level in order to reduce the computational load.

The motivation for applying the GP on residual from model results instead of the simulation data. Our simulation data for both crops and climate are high-dimensional gridded global data that cannot be modelled directly by GP. Several approximation techniques have been developed to handle multi-dimensional data. For instance, a low-rank approximation of the Gram matrix was described in [Rasmussen & Williams \(2006\)](#), [Drineas & Mahoney \(2005\)](#). The use of subsets of regressors and subsets of data to reduce the computational burden of the GP inversion for a large data matrix was investigated by [Silverman \(1985\)](#), [Wahba \(1990\)](#), [Wahba et al. \(1995\)](#). The following studies used Bayesian techniques for emulating high-dimensional data [Rougier \(2008\)](#), [Rougier et al. \(2009\)](#), [Heitmann et al. \(2006\)](#), [Higdon et al. \(2008\)](#), [Dancik \(2011\)](#). In particular, [Borinn & Zidek \(2012\)](#), [Finley et al. \(2011\)](#) used Bayesian methods for predicting crop yields. A good emulator must be as fast as possible to allow for relatively quick evaluation of many scenarios which are not feasible with these techniques. Therefore, the Bayesian procedures used in the studies mentioned above can not be directly applied to handle large dimensional and high-resolution crop data in this study.

Secondly, climate varies considerably from one spatial scale to another. This variability is determined largely by atmospheric circulation and its interactions with large-scale ocean currents. Regional or local climate is much more variable than global climate because climate on a large spatial scale is less influenced by internal dynamics of the continental or global climate. Therefore, even if we decide to use a PC decomposition or aggregation of the crop yield data (to reduce the dimension). We may not be able to apply that data reduction strategy to climate input parameters as they vary from one location to another. The decomposition of the data is non-trivial as climate may vary a lot within a country so that the average climate across a country may be a poor approximation to the local climates within the country. In addition LPJmL data is a non-linear response so large scale average climate may be an inappropriate input to our problem.

Thirdly, the LPJmL data that we are using are monthly data and they have been averaged to decadal level. We have already lost some information due to decadal averaging because the LPJmL model itself works on a daily time step. We do not want to lose further vital information from the data by further averaging or by applying principal components to reduce the dimension of our data. Applying PCA to reduce dimension of LPJmL will cause significant loss of information at early stage of the modelling. In addition, the resulting PC components will ignore information about the climate input variables thereby giving rise to PC components that might be difficult to explain by the original input variables. In order to proceed, we decided to emulate the high-dimensional data at a high spatial resolution using stepwise regression. We project the predictions and residuals obtained from stepwise regression results to a lower dimension by aggregating spatially to country levels. Then apply a GP regression for residuals interpolation from the new projected axes.

Another way to resolve the problem is to treat each spatial point as a GP and then emulate each grid cell individually a technique implemented in [Lee et al. \(2013\)](#) and [Lee et al. \(2012\)](#). However, this option is intractable since we have 59199 grid points meaning that about 236,796 emulators have to be constructed and validated for the five crops we are emulating. The procedure is described below.

### 3.2 Emulation procedure

The emulators were constructed in two stages. Our objective is to produce a statistical approximation that will link the model climate inputs  $\mathbf{X}$  to the deterministic vector outputs  $\mathbf{y}$  from LPJmL. Let the relationship between the climate input data from ClimGen and each output from LPJmL be represented by a model

$$y = f(\mathbf{X}) + \varepsilon = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \beta_1 x_{1,1}^2 + \dots + \beta_p x_{p,p}^2 + \beta_{1,2} x_1 x_2 + \dots + \beta_{(p-1),p} x_{(p-1)} x_p + \varepsilon \quad (1)$$

where  $\mathbf{y}$  is the (vector) simulated mean decadal change for LPJmL crop-yield over the grid cells for all scenarios,  $\varepsilon$  represents all the unexplained variation in  $f(\mathbf{X})$  which is not captured by the stepwise regression method in stage 1. Values of  $\varepsilon$  are spatially correlated for scenarios that are very close in the input space. In the second stage, we model  $\varepsilon$  as a Gaussian random process with unknown mean and covariance functions.

GP is an extension of multivariate Gaussian distributions to infinite dimensionality with associated mean and covariance matrix.

### 3.2.1 First stage

The LPJmL crop-yield data were based on simulations for 59199 grid cells on  $0.5^\circ$  by  $0.5^\circ$ , but here we only focused on current crop growing areas (the other cells are truncated). We built separate emulators for the rainfed and irrigated crops. Four crops were selected for emulation; temperate cereal, rice, maize and oil crop. The oil crop is a maximum yield among soybean, sunflower and rapeseed. The average decadal yield given by LPJmL from 2005-2095 was computed for each crop and each scenario. We then obtained the change in yields relative to the baseline values. Baseline values are the average decadal yield of LPJmL outputs and climate data for the period 2005-2014, and are used as a reference for computing the relative change in yields and climate. We calculated the change in seasonal climate variables for the input variables listed in Table ?? . Baseline yields, baseline seasonal climate and change in seasonal climate variables are used as input to the first stage emulator.  $\text{CO}_2$ , latitude, soil type and  $\text{LAI}_{\max}$  are also included as additional inputs. We produce different emulators for irrigated and rainfed crops but combine all the scenario together.

The emulators were built from two GCMs with relatively moderate equilibrium climate sensitivity, CCCMA-CGCM31 and CCSR-MIROC32HI, four RCPs, and two simulation categories (with and without  $\text{CO}_2$  fertilization effect), giving 16 different scenarios. Each scenario has seven crop management levels and eight time-slices, with each time-slice consisting of 16 649, 6913, 19 139 and 8427 valid grid points per scenario for rainfed temperate cereal, rice, maize and oil, respectively, and 8963, 7325, 9146 and 2731 for irrigated crops. Another five GCMs, UKMO-HADGEM1, GISS-MODELER, GISS-MODELEH, IPSL-CM4 and CCSR-MIROC32MED were selected for cross-validation purposes.

We use observations from 5000 randomly sampled grid points because the stepwise algorithm could not handle the whole dataset. The 5000 grid points are fixed across the time-slice, RCP, GCM and management levels as well as simulation categories and similarly in each of 37 input variables (see Table ??) in Appendix. Irrigated oil has less than 5000 observations so we used all the simulation data. We then fitted a stepwise regression model (including linear, quadratic and all two-way interactions) to the sampled data for each change in crop-yield when crops are rainfed. This procedure was carried out for each of the four crops and repeated for the case where crops are irrigated. Further details of this procedure are given in (Oyebamiji et al., 2015).

### 3.2.2 Second stage

Consider a single crop/irrigation regime/management level/time slice combination and let  $\mathbf{y}_i$  be the vector of changes in yield given by LPJmL for that combination in the  $i^{\text{th}}$  scenario ( $i = 1, \dots, 16$ ). We let  $\tilde{\mathbf{y}}_i$  be the corresponding predictions given by the stage 1 emulator and  $\boldsymbol{\varepsilon}_i = \mathbf{y}_i - \tilde{\mathbf{y}}_i$  is the error in prediction. Each  $\tilde{\mathbf{y}}_i$  and  $\boldsymbol{\varepsilon}_i$  is an  $N \times 1$  vector, where  $N$  denotes the number of grid cells for that crop/irrigation regime. As  $\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_{16}$  are predictions on *training* data, the values of  $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_{16}$  are known.

Given a new vector of predictions,  $\tilde{\mathbf{y}}^{\text{new}}$ , from a new scenario where the emulator values are unknown, the aim is to estimate the error of  $\tilde{\mathbf{y}}^*$  from  $\mathbf{E} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_{16})$  and  $\tilde{\mathbf{Y}} = (\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_{16})$ . In order to reduce the dimension of data we apply a PC decomposition to the  $16 \times N$  matrix  $\tilde{\mathbf{Y}}^T$ , and select just the first four principal components. The resulting  $16 \times 4$  matrix of PC scores given by these four components is denoted as  $\tilde{\mathbf{U}}$ . The columns of matrix  $\tilde{\mathbf{U}}$  will be used as explanatory variables for the GP regression of residual patterns. The procedure is described briefly here.

Let  $\tilde{\mathbf{Y}}^T$  be the predictions given by stage 1 for the 16 training scenarios. Matrix  $\mathbf{U}$  is obtained by projecting  $\tilde{\mathbf{Y}}^T$  onto new axes using a PC decomposition. Apply eigendecomposition to matrix  $\tilde{\mathbf{Y}}^T$  gives

$$\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T = \Gamma\Lambda\Gamma^T \quad (2)$$



where  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$  is a diagonal matrix of eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{16}$ , and  $\Gamma = (\gamma_1, \dots, \gamma_{16})$  is an  $16 \times 16$  orthogonal matrix, where  $\gamma_i$  is the eigenvector corresponding to  $\lambda_i$ . The PC transformation from  $\tilde{\mathbf{y}}$  to  $\tilde{\mathbf{u}}$  is given by

$$\tilde{\mathbf{y}} \rightarrow \Gamma^T \tilde{\mathbf{y}} = \tilde{\mathbf{u}}. \quad (3)$$

Put  $\tilde{\mathbf{u}}_i = \Gamma^T \tilde{\mathbf{y}}_i$  for  $i = 1, \dots, 16$  and put  $\tilde{\mathbf{u}}^{new} = \Gamma^T \tilde{\mathbf{y}}^{new}$ , where  $\tilde{\mathbf{y}}^{new}$  are the stage 1 predictions for any fresh scenario, where  $u_j^{new}$  and  $u_{ij}$  are the  $j^{th}$  components of  $\tilde{\mathbf{u}}^{new}$  and  $\tilde{\mathbf{u}}_i$ , respectively. We used 4 PC, these explained at least 95% of the variation in  $\tilde{\mathbf{Y}}$ . Let  $\hat{\Gamma}$  be the first four columns of  $\Gamma$ , so  $\hat{\Gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ . We put  $\tilde{\mathbf{U}} = \tilde{\mathbf{Y}}^T \hat{\Gamma}$ . Each column of  $\tilde{\mathbf{U}}$  is now linearly uncorrelated and matrix  $\tilde{\mathbf{U}}_{16 \times 4}$  is a required explanatory variable for the GP regression. Let  $\epsilon_i = \mathbf{y}_i - \tilde{\mathbf{y}}_i$  as defined above.

A GP regression is used to estimate  $\epsilon^{new}$ , the corresponding error in  $\tilde{\mathbf{y}}^{new}$ . A separate GP regression is used for each component of  $\epsilon^{new}$ . The dimension of the residual matrix  $\mathbf{E}^T$  is  $16 \times N$ , ( $N = 16648$  for rainfed cereal) which is very large for the purposes of GP regression and falls under multivariate outputs category which is computationally intensive. We do not want to emulate the linear combinations by applying PCA to reduce the dimension as we did for the explanatory variable. This is because the PCA approach will increase the computational burden. In addition, it can be difficult to back-transform the results to the original scale for proper interpretation. Since columns of  $\mathbf{E}^T$  represent global spatial points, we apply spatial aggregation to reduce the dimension to a country level using equation (13). We took the outputs as a combination of separate univariate output and modelled each independently.

The spatial aggregation will transform  $\mathbf{E}^T$  to matrix  $\mathbf{Z}$  of dimension  $16 \times 186$  ( $\mathbf{E}_{16 \times N}^T \rightarrow \mathbf{Z}_{16 \times 186}$ ). Details of the aggregation are provided under section 3.3. So for the  $n^{th}$  GP regression (the  $n^{th}$  country) the values of the dependent variable are  $\mathbf{z}$ , where  $\mathbf{z}$  is the  $n^{th}$  column of  $\mathbf{Z}$  (ie  $\mathbf{z} = \mathbf{Z}[:, n]$ ,  $\mathbf{z} = (z_1, \dots, z_{16})$ ) is a vector of observed residual for  $n^{th}$  country.

After aggregating the residual to a country level, we take one country at a time and form a separate GP regression equation for that country. The data for one of these regressions is the  $(16 \times 1)$  vector of responses  $\mathbf{z}$  (the residuals for that country) and the  $(16 \times 4)$  matrix  $\tilde{\mathbf{U}}$ , which holds the values taken by the explanatory variables.

For the new scenario, put  $\mathbf{u}^{new} = \hat{\Gamma}^T \tilde{\mathbf{y}}^{new}$ . Then the estimate of the prediction error in  $\tilde{\mathbf{y}}^{new}$  at the  $n^{th}$  country is derived as following. Set up the Gaussian process by selecting a mean and a covariance function. Then the residual  $\mathbf{z}$  can be modelled as a GP regression given by

$$\mathbf{z} = g(\mathbf{u}) = \mathbf{h}^T(\mathbf{u})\beta + \delta(\mathbf{u}), \quad (4)$$

where  $\mathbf{h}(\mathbf{u})$  is a vector of regression functions and is chosen to reflect the functional form of our response data. Under GP regression,  $\mathbf{h}(\cdot)$  can have either a constant mean function ( $\mathbf{h}(\cdot) = 1$ , corresponding to only intercept term) or can be modelled as a simple linear regression function of the inputs ( $\mathbf{h}(\mathbf{u}) = (1, \mathbf{u}^T)$  including intercept term). We compare both cases with relatively little difference in this paper.

$\beta$  is an unknown hyperparameter to be estimated,  $\delta(\cdot)$  is a stationary GP representing stochastic noise with mean zero and covariance function

$$K = \text{Cov}(\mathbf{z}(\mathbf{u}), \mathbf{z}(\mathbf{u}^T)) = \sigma^2 C(\mathbf{u}, \mathbf{u}^T),$$

where  $C(\mathbf{u}, \mathbf{u}^T)$  is a correlation function with an hyperparameter  $\alpha$ . The covariance function  $K$  must be a semi-positive definite function to ensure that it can be inverted; therefore we choose a squared exponential correlation function which is infinitely differentiable. In addition, this correlation function  $C(\mathbf{u}, \mathbf{u}^T)$  decreases as the distance between  $\mathbf{u}$  and  $\mathbf{u}^T$  increases, and usually has the form

$$\mathbf{C} = \text{Cor}(\mathbf{u}, \mathbf{u}^T) = \exp \left[ \sum_{j=1}^4 -\alpha_j (\mathbf{u}_j - \mathbf{u}_j^T)^2 \right]. \quad (5)$$

where  $\alpha = (\alpha_1, \dots, \alpha_4)$  and it denotes the smoothness parameter that measure the rate of changing of response as input value changes, and it will be estimated from the training data.

The Gaussian process has the same functional form as the multivariate Gaussian distribution, and it is denoted by its mean and covariance functions. Therefore, if we assume a Gaussian prior process for the output function  $g(\cdot)$  and update this with our training data  $\mathbf{D}^T = [(z_i, g(u_{ij})); i = 1, \dots, 16; j = 1, \dots, 4]$ . This computation would produce a posterior distribution which is also a Gaussian distribution. GP regression will be used to estimate  $z^{new}$ , the corresponding error of  $\mathbf{u}^{new}$ . In order to predict the response at new input point  $\mathbf{u}^{new}$ , given some training data  $\mathbf{D}$ , the joint distribution of the observed values  $\mathbf{D}$  and test point  $z^{new}$  can be obtained using a multivariate Gaussian identity. Suppose that

$$z^{new}|\mathbf{D} \sim N \left[ \begin{pmatrix} h^T \\ \mathbf{H} \end{pmatrix} \beta, \begin{pmatrix} \mathbf{c}(\mathbf{u}) & \mathbf{t}(\mathbf{u}^T) \\ \mathbf{t}(\mathbf{u}) & \mathbf{C} \end{pmatrix} \sigma^2 \right] \quad (6)$$

where  $\mathbf{t}(\mathbf{u}) = \text{Cor}(\mathbf{u}, \mathbf{u}^{new})^T$  is a correlation vector of training data  $\tilde{\mathbf{U}}$  with a new input point  $\mathbf{u}^{new}$ ,  $\mathbf{C}$  is a  $16 \times 16$  correlation matrix among the training data  $\tilde{\mathbf{U}}$  (ie design matrix) and,  $\mathbf{c} = \text{Cor}(\mathbf{u}^{new}, (\mathbf{u}^{new})^T)$  is the correlation between the test data. Therefore, the conditional posterior distribution is given by the form

$$P(z^{new}|\mathbf{D}, \beta, \sigma^2) \sim N(\mu^\bullet, \mathbf{K}^\bullet) \quad (7)$$

and after some algebraic manipulation (by integrating out the hyperparameters) the posterior mean and covariance functions are given respectively as

$$\mu^\bullet(\mathbf{u}) = h^T(\mathbf{u})\hat{\beta} + \mathbf{t}(\mathbf{u})\mathbf{C}^{-1}[\mathbf{D} - \mathbf{H}(\mathbf{u})\hat{\beta}] \quad (8)$$

$$\mathbf{K}^\bullet(\mathbf{u}, \mathbf{u}^T) = \hat{\sigma}^2 \left( \mathbf{c}(\mathbf{u}) - \mathbf{t}(\mathbf{u}^T)\mathbf{C}^{-1}\mathbf{t}(\mathbf{u}) \right). \quad (9)$$

which are directly follow from subsection ?? in Appendix 1, where  $\mathbf{H}$  is the  $16 \times 4$  matrix given as  $\mathbf{H}^T = [\mathbf{h}(\mathbf{u}_1), \dots, \mathbf{h}(\mathbf{u}_{16})]$ , with  $\mathbf{u}_i$  denoting the  $i^{th}$  design point and hyperparameters  $\beta$ ,  $\sigma^2$  and  $\alpha$  are obtain using the MLE method. The mean function  $\mu^\bullet(\mathbf{u})$  is an estimate of function  $g(\mathbf{u})$  and the covariance function  $\mathbf{K}^\bullet(\mathbf{u}, \mathbf{u}^T)$  is a measure of uncertainty associated with this mean function.

$$\hat{\beta} = (\mathbf{H}\mathbf{C}^{-1}\mathbf{H})^T\mathbf{H}\mathbf{C}^{-1}\mathbf{z} \quad (10)$$

$$\hat{\sigma}^2 = \frac{1}{n} \left[ (\mathbf{z} - \mathbf{H}\hat{\beta})^T \mathbf{C}^{-1} (\mathbf{z} - \mathbf{H}\hat{\beta}) \right]. \quad (11)$$

See further derivation of GP parameters in the Appendix ?? . Having obtained both the posterior mean and variance estimates of the actual residual at a new input point for each country, we need to compute the final predictions from the emulator, which are then given as  $\mathbf{y}^{new}$

$$\hat{\mathbf{y}}^{new} = \hat{\mathbf{y}}^{new} + \mu^\bullet(\mathbf{u}) \quad (12)$$

However, note that our  $\hat{\mathbf{y}}^{new}$  is half degree data thus it has to be aggregated to have the same spatial resolution as the residual. Therefore, we follow a similar procedure using equation 13 for its transformation. We used equation (8) to estimate the prediction error separately for each country and equation (12) is the revised estimate for one country. To combine the calculations for all countries into a single step, we fitted an independent GP regression in parallel to the columns of  $\mathbf{Z}$ .

We applied the same concept to model the other crops (rice, maize and oil). This process is repeated for all the time slices, RCPs and management levels. In order to test the performance of the emulator, a complete cross-validation is performed using another five selected GCMs, UKMO-HADGEM1, GISS-MODELER,



GISS-MODELEH, IPSL-CM4 and CCSR-MIROCMED. Some cross-validation results are provided in section 4.

### 3.3 Spatial aggregation procedure

In order to apply the GP regression for residual interpolation, we reduce the resolution of the residuals data from half degree to a country level by aggregating the residual matrix  $\mathbf{E}^T$ . For the evaluation of GP regression of residual, country-level residual values are computed by multiplying the residual at each grid cell by the crop growing area for each cell and separately for rainfed and irrigated crops. We calculated the area-weighted sum for each country to determine the total residual for that country, and finally dividing this total residual by the total (country) area.

Therefore, for any  $i^{th}$  residual scenario,

$$\mathbf{Z}_{[i,n]} = \frac{\sum_{j=1}^m \mathbf{E}_{[i,j]}^T \times \mathbf{A}_j}{\sum_{j=1}^m \mathbf{A}_j} \quad (13)$$

where  $m$  is the number of grid cells that fall into country  $n$ ,  $\mathbf{E}$  is the residual as defined earlier and  $\mathbf{A}_j$  is the crop growing area for the  $j^{th}$  grid cell and  $i = (1, \dots, 16)$ ,  $j = (1, \dots, N)$ .

## 4 Results

We choose one combination of crop/management/time point from a particular scenario to demonstrate the result. The results shown here are for the 8<sup>th</sup> decade corresponding to change between the average (2085-2094) and baseline period average (2005-2014) with CO<sub>2</sub> fertilization effect. The climate at this time point is assumed to be characterized by a relatively large climate uncertainty associated with a high impact on yield change, and RCP 6 which is a moderate emission scenario under management level 5. Crop management level is a measure of vegetation density in cropping systems as influenced by machinery, fertilizer application. A well-managed system is assumed to have values of  $\geq 5$  for developed countries while under-developed countries like Sub-Sahara Africa are assigned values between 1 and 3. Figure 1 shows the percentage of variance explained by each PC component for the four crops we considered in this paper. We can see that the first 4 PCs explained at least 95% of the variance for each crop.

Figure 2 is the GP emulation results for rainfed temperate cereal under RCP6, management level 5 and 8<sup>th</sup> decadal change for UKMO-HADGEM1. Figure 2a shows the performance of the 106 GP rainfed emulators for the cereal. Each GP corresponds to individual countries where the rainfed cereal crop is grown.

The four countries with the largest change in yield are Tajikistan, Montenegro, Canada and Luxembourg see Table (3) for a list of UN countries. Uruguay and Morocco are two countries with a large reduction in cereal yield. There is relatively little change in other countries. The observed values for LPJmL are well captured within the 95% C.I of the emulator predictions. The emulator relatively well predicts Tajikistan with the largest change. Uruguay has the lowest change in yield and is under-predicted by emulator, with values  $-1.21 \pm 0.05$ ,  $-1.46$  ton/Ha given respectively by emulator and LPJmL.

Figure 2b is a bilinear plot between LPJmL and the emulator for 106 countries. The LPJmL values are arranged in ascending order from least to highest. The plot shows more clearly that the change in yield for LPJmL are clustered together for most of the countries and are within 95% C.I limit. The two end points correspond to Tajikistan and Uruguay, and uncertainties at either end are large as seen in the plot. The total percentage of variance explained by emulator for cereal is 92%.

Figure 3a shows that the observed yield (LPJmL) and predicted (emulator) are mostly within 95% C.I. Overall, we see that the cereal is well predicted for both rainfed and irrigated yields with majority of points lie within their uncertainty limits. Further test involves computing cross-validated root mean squared error

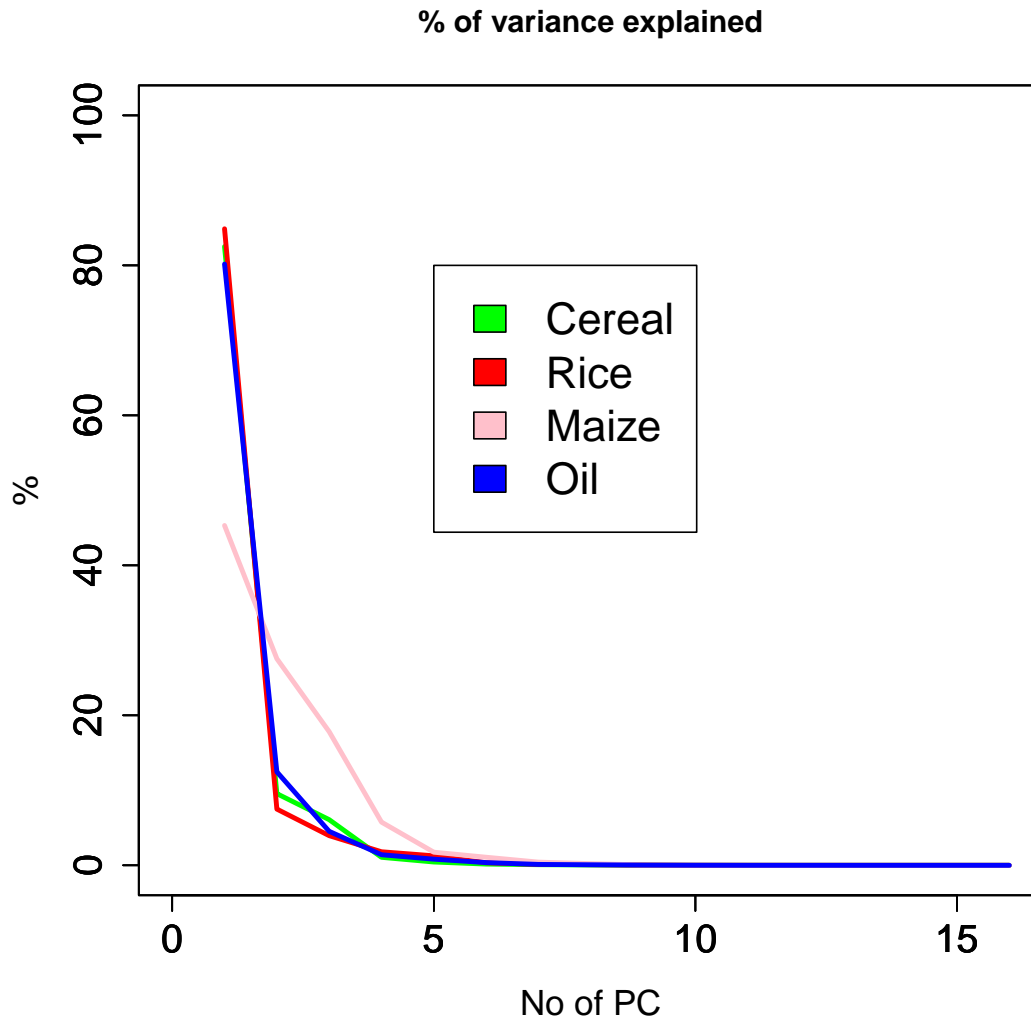
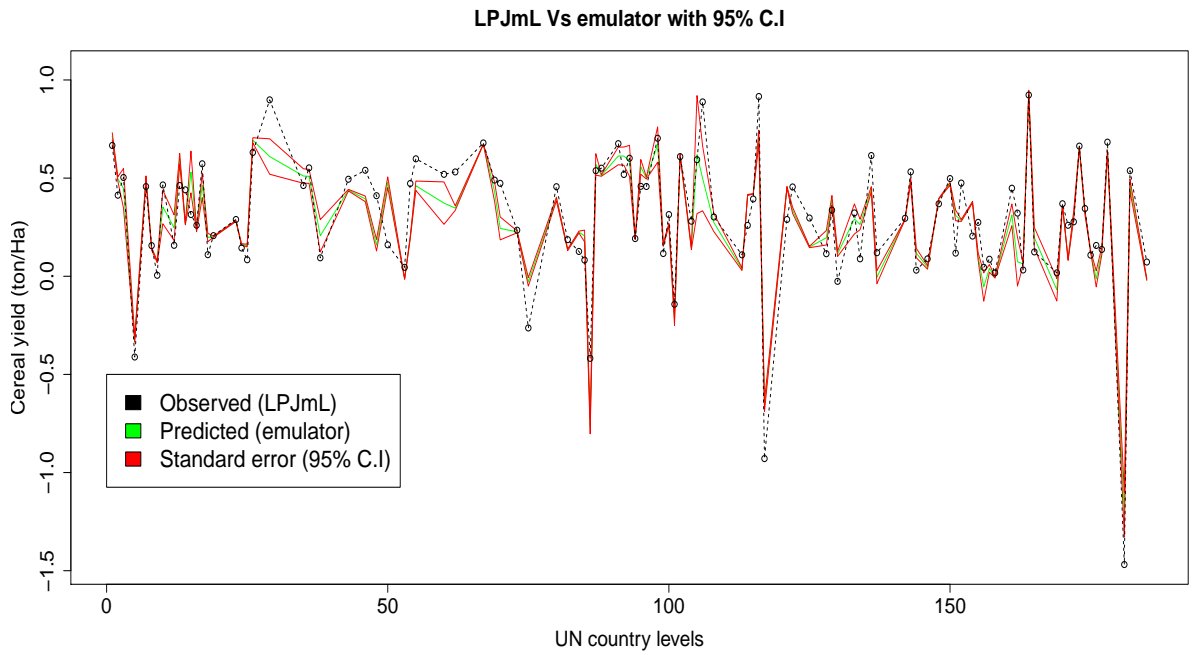


Fig. 1: % of the variance explained by each PC for the four rainfed crops. The plots correspond to crop-yield for management level of 5, RCP6 with CO<sub>2</sub> fertilization from UKMO-HADGEM1.

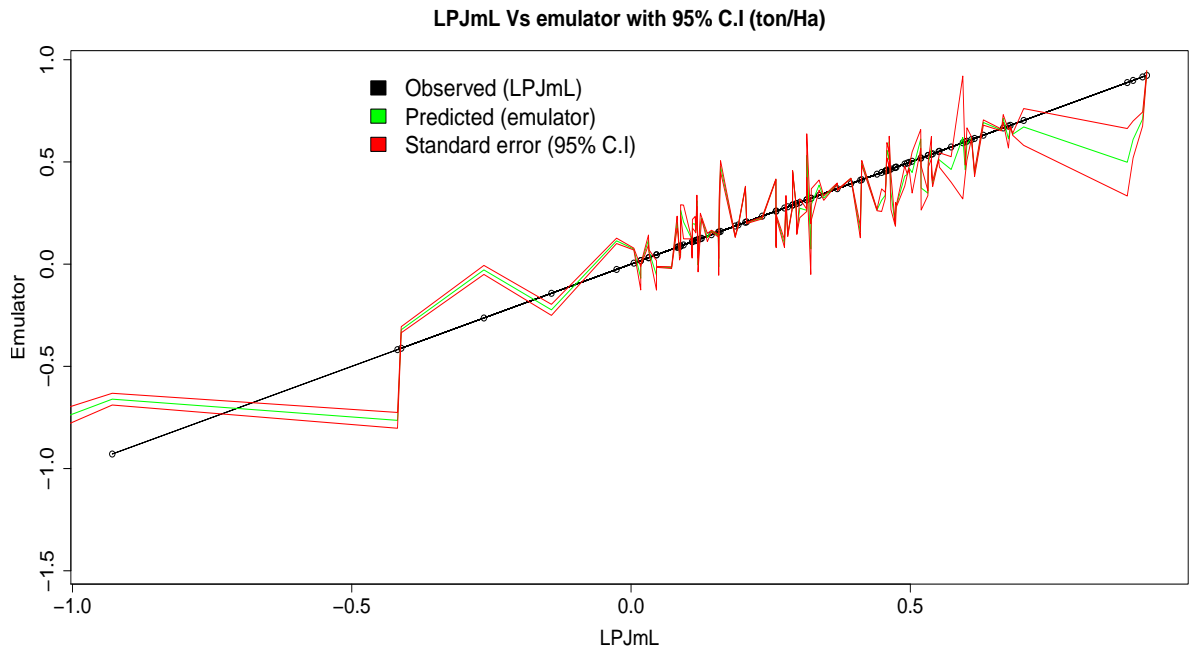
( $RMSE_{CV}$ ) is given by equation 31. The values are 0.0082 and 0.0018 ton/Ha for rainfed and irrigated respectively. We can clearly see that predictions are much better (low  $RMSE_{CV}$ ) for irrigated than for rainfed yields.

We show results for the major producing countries. In Figure 5a, emulator predictions are similar to LPJmL values except for India where the emulator under-predicts the change in yield. LPJmL produces about 17% while emulator gives  $13 \pm 0.4\%$  for India. There is a substantial change in yield in the USA than for other major producers of cereal crop with both LPJmL and emulator estimate  $\geq 33\%$ . For the irrigated cereal in Figure 5b, emulator predicts relatively well the change in yield for China, India and Russia. In each of this case, the LPJmL values are within the 95% C.I band. Emulator under-predict the change in USA. There is a general increase in yield under climate change in these countries. The growth varies between 6-36% across the country for both rainfed and irrigated crops.

Consider Figures 5c and Figure 5d, China has a substantial change with a rise of 26% in yield for rainfed and a growth of 30% under irrigation. The emulator predicts  $28 \pm 0.18\%$  and  $30 \pm 2.2\%$  respectively. The results of other countries are relatively close except under rainfed in Bangladesh where the emulator under-predicts the yield. For irrigated rice, emulator under-predicts change in Indonesia and over-predicts the change for Bangladesh. In Figure 6a, maize exhibits a slight increase in yield with the globally averaged yield change increasing by 2% in China and Brazil. In contrast, there is a relatively steady reduction in yield in USA and



(a) LPJmL and emulator predictions including their 95% C.I vs. UN country levels

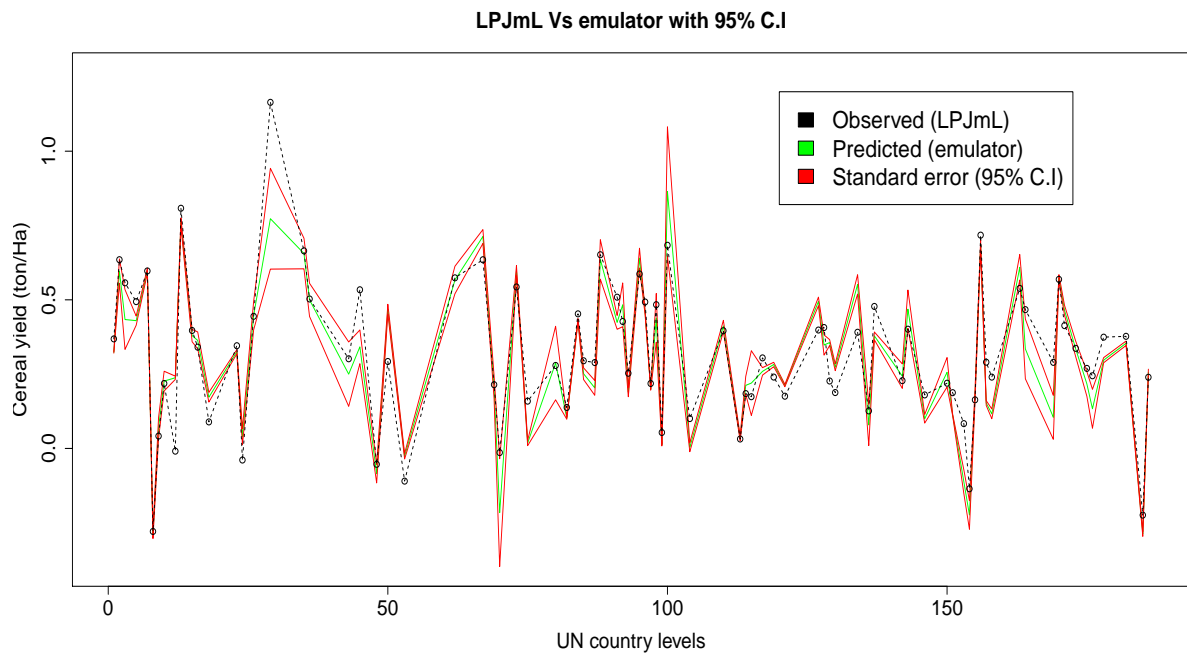


(b) Pair plot of emulator predictions vs. LPJmL values including their 95% C.I

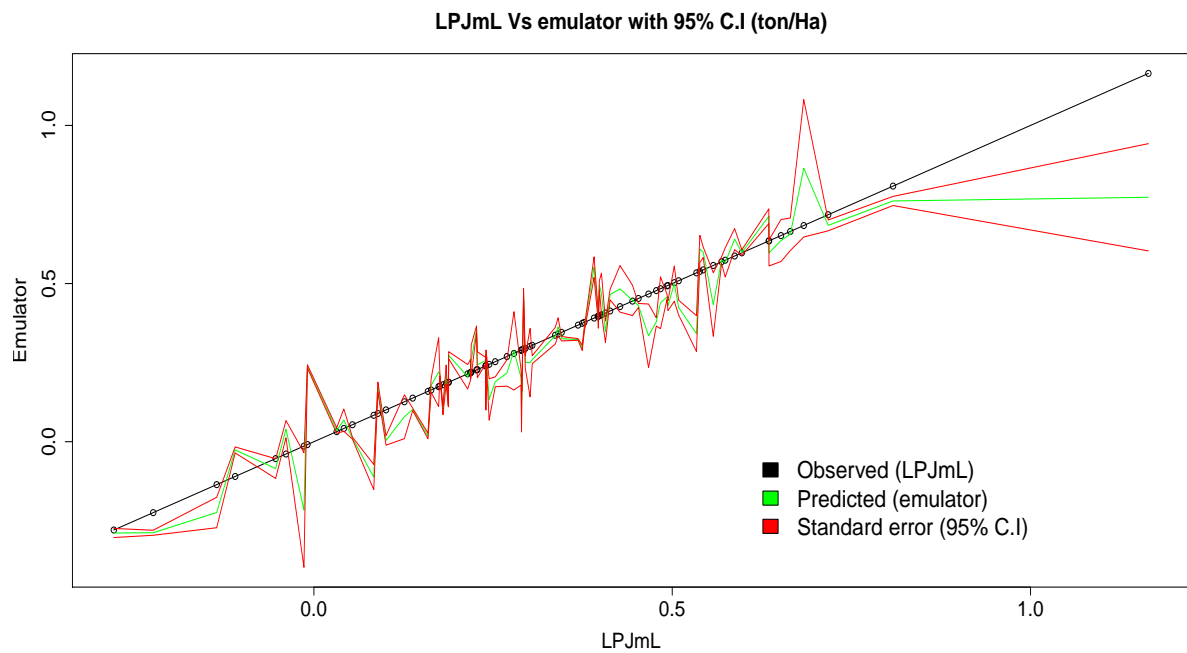
Fig. 2: Cross-validation for rainfed cereal; LPJmL and emulator with its 95% C.I for some countries levels.

France as simulated by LPJmL. The emulator, on the other hand, seems to be badly behaved as the values are poorly predicted by the emulator. For instance, while the LPJmL produces a growth of 2.5% in Brazil, emulator projects a reduction of  $1.2 \pm 0.12\%$ . In France, LPJmL gives a decline of 1.5% emulator predicts a rise of  $4 \pm 1\%$ .

Emulator predictions under irrigation in Figure 6b are much better than for rainfed maize. LPJmL values are within 95% confidence band. There is a slight increment in yield of less than 3% in China and USA while France has a considerable large growth of 20%. However, there is a decline in maize yield in Brazil. Figure 6c



(a) LPJmL and emulator predictions including their 95% C.I vs UN country levels.



(b) Pair plot of emulator predictions vs LPJmL values including their 95% C.I

Fig. 3: Cross-validation for irrigated cereal yield; LPJmL and emulator with its 95% C.I for some countries levels.

shows a large increment of more than 60% in Argentina oil and Figure 6d indicates a modest increase in oil yield in Brazil. The emulator predicts better result in comparison with LPJmL values.

Overall, there is an increase in yield for most of the countries as expected because of the beneficial effect of CO<sub>2</sub> fertilization included in this simulation. However, maize shows a reduction in yield in USA, France and Brazil.

We now consider the spatial map of rainfed cereal in Figure 4 (top) in more detail. The white coloration in this map represents regions that are not currently growing cereal crop. Most of the countries are reasonably well-predicted. The Russia Republic and Canada are under-predicted by the emulator. For instance, while the

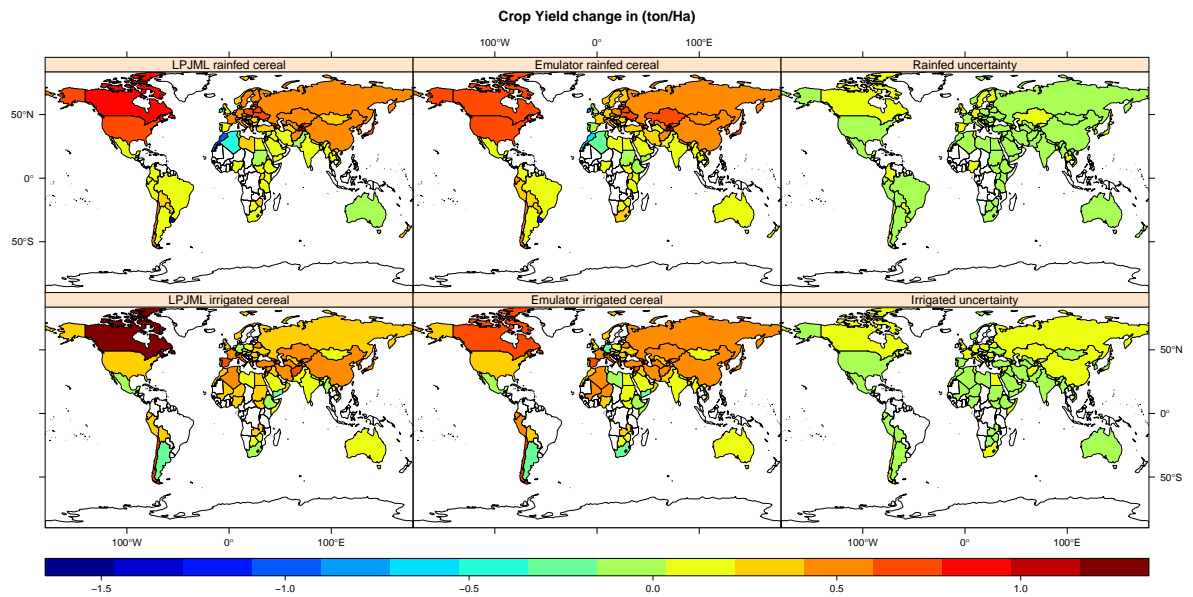


Fig. 4: Cross-validation of the comparison between LPJmL and LPJmLem for rainfed and irrigated temperate cereals, plotted as mean decadal change in yield between (2085-2094) and (2005-2014) with their 95% C.I. The plots correspond to crop-yield for management level of 5, RCP6 with CO<sub>2</sub> fertilization from UKMO-HADGEM1.

LPJmL simulates increase in yield of 0.52 and 0.90 ton/Ha for the Russia Federation and Canada, emulator predicts growth of  $0.49 \pm 0.01$  and  $0.61 \pm 0.09$  ton/Ha respectively. Predicted values by the emulator are higher than for the LPJmL values for Mongolia, Tajikistan and South Africa. For instance, while LPJmL gives increase in yield of 0.20 ton/Ha emulator produces  $0.38 \pm 0.003$  ton/Ha for South Africa.

Emulators predict increases in yield, as does LPJmL for most of the countries under this scenario. This increase is expected because of beneficial CO<sub>2</sub> effect. There are significant yield changes for Europe, China, USA, Russia, and these countries have in common a well-managed land system. In addition, increasing in warming conditions, as projected under climate change will give rise to prolonged growing season in these regions that will cause an increase in yield (Bates et al., 2008, IPCC, 2013, Zhou et al., 2013). In particular, Wu et al. (2014) associated significant increases in food production in China to technological advancement and changing in agronomic practices in that country.

Yield changes are significantly higher for high latitude and middle than for low latitudes in agreement with some literatures Rosenzweig et al. (1994), IPCC (2007), Parry et al. (2005). Climate change will cause low latitude to experience a greater degree of heat and water stress which will cause a decline in yield even in the presence of the CO<sub>2</sub> fertilization effect. For instance, in Australia there is a fairly little change in yield that could probably due to a large variability in precipitation, extended droughts and a greater incidence of extreme weather events in this region which could disrupt agricultural production. In addition, Australia and New Zealand exhibit a wide diversity of climates with a significant constraint on water resources as described in (IPCC, 2013). Algeria, Ireland, Morocco and Uruguay also have a substantial reduction in yield. Uncertainty levels are low and vary from one country to another because each country is modelled by an independent Gaussian process regression.

Under irrigation, Figure 4 (bottom) yield change are relatively lower in USA, China and Russia compare to rainfed system unlike Canada that has a larger increase yield than for rainfed. Reduction in crop yield is large in Argentina, Yemen, South Africa and Ethiopia. Increase in yield changes is substantial for Canada, Armenia and Spain. Similar to rainfed cereal, lower yield changes are associated with medium and low latitude countries than for high latitude countries.

Next, we extend our cross-validation to other crops. Figure 7 (top) is for rainfed rice. The plot indicates that emulators predict the change in yield reasonably well in most countries. The emulator under-predicts the change in yield in Russia Federation and Argentina. Both LPJmL and emulator produce a higher increase in

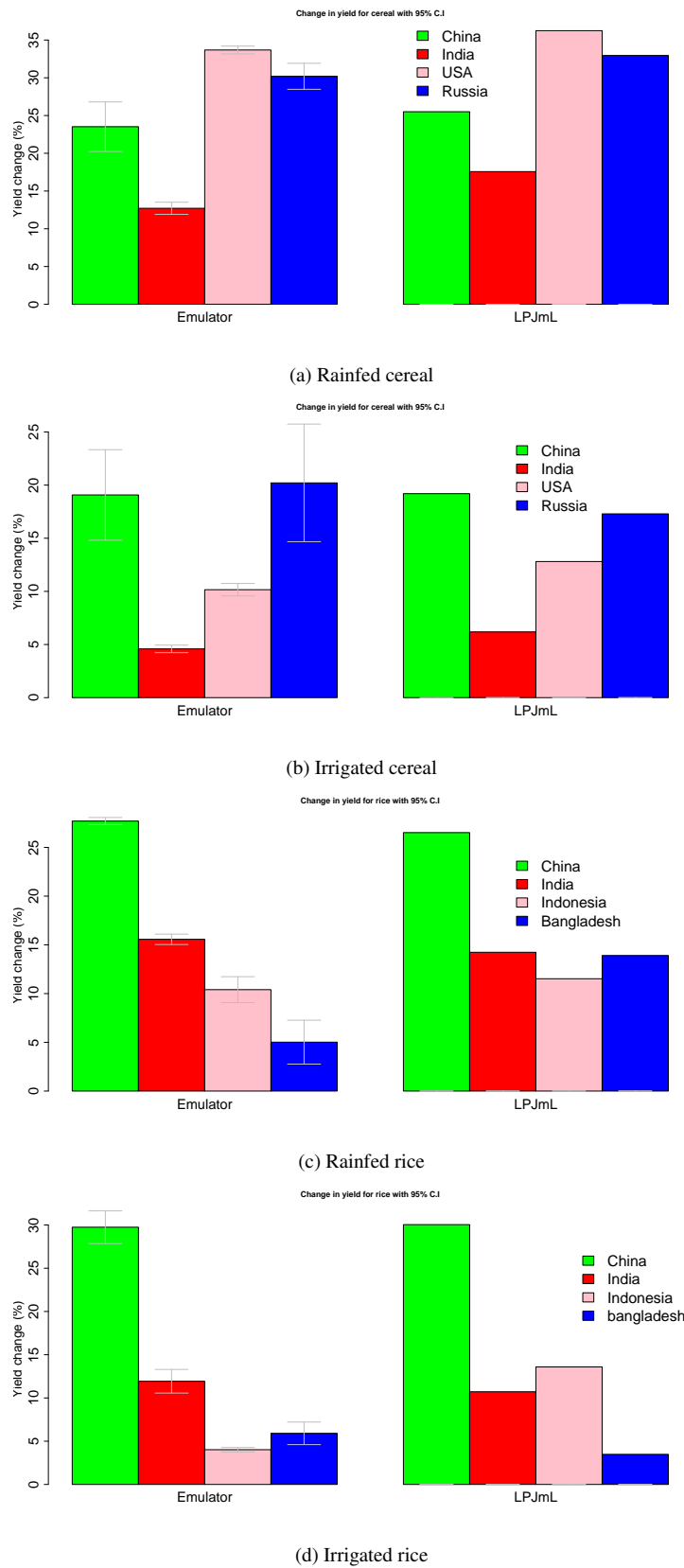


Fig. 5: Change in percentage yield by emulator with 95% C.I. in comparison with LPJmL values for major producers of selected crops.

yield for Russia Federation and Argentina than for the USA and China. Uncertainties around predictions in most countries are also quite low.



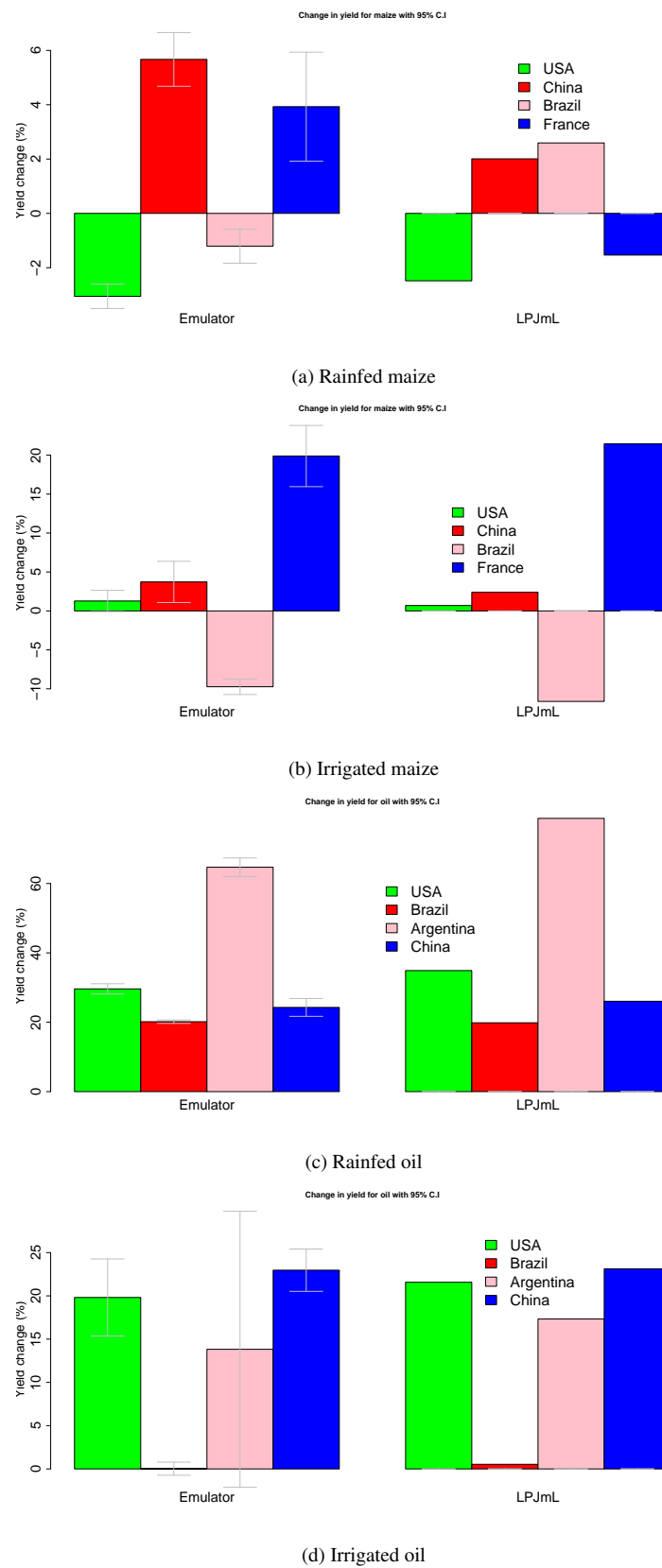


Fig. 6: Change in percentage yield by emulator with 95% C.I. in comparison with LPJmL values for major producers of selected crops.

For irrigated rice (bottom) plot, LPJmL values are quite well predicted in most countries. Chile, Morocco and Macedonia including Hungary have a substantial increase in yield than for other countries. Uncertainty

levels are quite low for most of the predictions. Consider Figure 8 (top), predictions are quite close to the LPJmL values in most regions for both rainfed and irrigated. Uncertainties associated with predictions are relatively similar for most countries, and the values are small. The Russia Federation has a high value under rainfed rice as we have earlier observed, a similar high value is noticed under irrigated crop. Maize, unlike other crops, exhibits little change in yield under this scenario.

Figure 9 for rainfed oil shows a rise in yield for most countries, with emulator predictions comparable to the LPJmL simulation. There is an under-prediction in Mongolia. The results are also similar under irrigation system. Uncertainty levels are generally low.

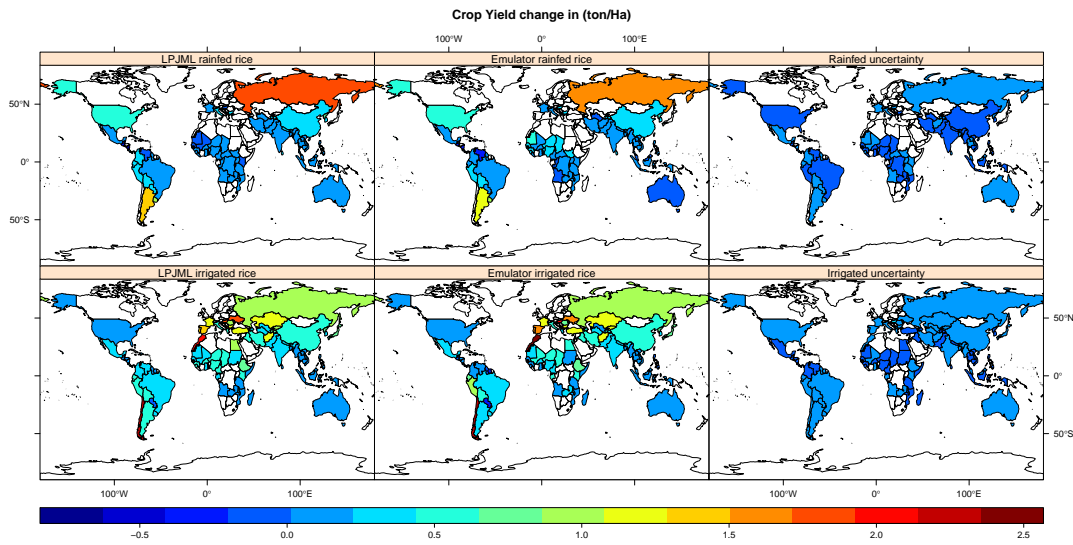


Fig. 7: Cross-validation of the comparison between LPJmL and LPJmLem for rainfed and irrigated temperate rice, plotted as mean decadal change in yield between (2085-2094) and (2005-2014) with their 95% C.I. The plots correspond to crop-yield for management level of 5, RCP6 with CO<sub>2</sub> fertilization from UKMO-HADGEM1.

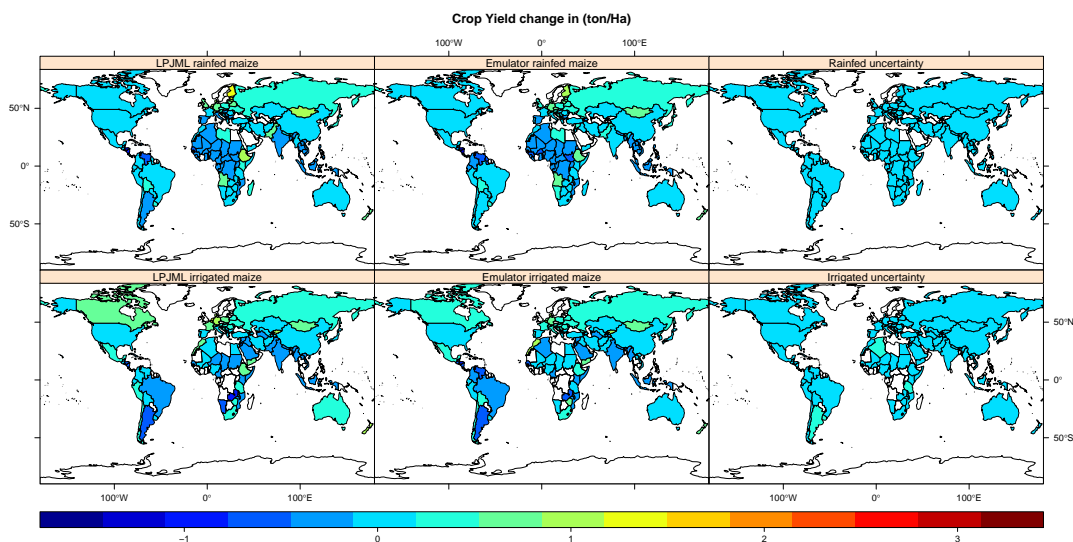


Fig. 8: Cross-validation of the comparison between LPJmL and LPJmLem for rainfed and irrigated temperate maize, plotted as mean decadal change in yield between (2085-2094) and (2005-2014) with their 95% C.I. The plots correspond to crop-yield for management level of 5, RCP6 with CO<sub>2</sub> fertilization from UKMO-HADGEM1.

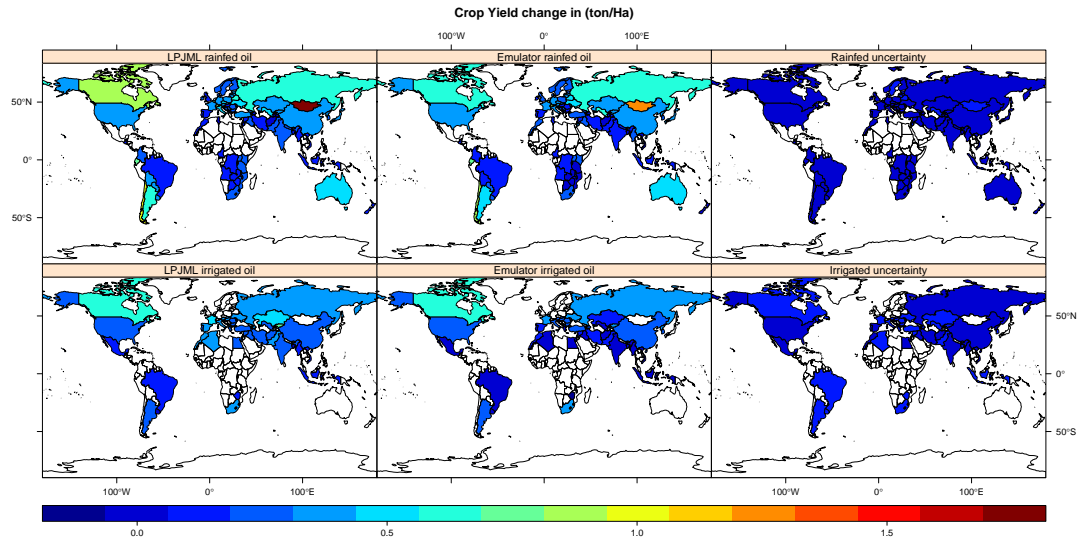


Fig. 9: Cross-validation of the comparison between LPJmL and LPJmLem for rainfed and irrigated temperate oil, plotted as mean decadal change in yield between (2085-2094) and (2005-2014) with their 95% C.I. The plots correspond to crop-yield for management level of 5, RCP6 with CO<sub>2</sub> fertilization from UKMO-HADGEM1.

We summarise the results of the emulator in Table 1 for all management levels, RCP and time point with CO<sub>2</sub> fertilization for UKMO-HADGEM1. We also compare the performance of the GP emulator with the WLS techniques. In this case, we use the same procedure set out in this paper but instead of performing GP regression we used the WLS approach. Proportion of variances explained is shown below. Table 1 indicate that GP regression is slightly better than the WLS method for all the cases except for rainfed oil. Irrigated cereal and maize results are much better under GP models than for WLS approach. However, the two techniques produce relatively similar results for irrigated rice and oil. GP emulator explains between 70-94% variance while WLS ranges from 63-93% for both rainfed and irrigated crops.

Consider cross-validation for four additional GCMs in Table 2. We see that CCSR-MIROCMED results are much better than for other GCMs as expected because a very similar GCM, CCSR-MIROC32HI, was part of the training data. Results from GP models are much better than WLS technique among all GCMs except for rice under IPSL-CM4 where the result is slightly better than for GP models.

Table 1: Table of cross-validated proportion of variance  $\rho$  showing the overall performance of the GP emulators, and in comparison with WLS method, for rainfed and irrigated crops, with all management levels, RCPs, and time slices, but with CO<sub>2</sub> fertilization only. The list of countries are shown in Table 3.

$\rho$	Number of countries		GP		WLS	
crops	rainfed	irrigated	rainfed	irrigated	rainfed	irrigated
cereal	106	83	0.89	0.94	0.81	0.87
rice	79	95	0.82	0.92	0.77	0.93
maize	141	100	0.70	0.80	0.63	0.76
oil <sub>max</sub>	139	64	0.89	0.88	0.91	0.88

Figures 10a average global changes in yield predicted by emulator in all countries and compare to LPJmL values. Emulator predicts global average change in yield for cereal and rice relatively well with 14% and 13% respectively with LPJmL values within 95% C.I. Maize and oil crops are under-predicted by the emulator. Predicted maize values from the emulator is much lower to the LPJmL values. LPJmL simulates a small increase of 1.9% for maize; emulator predicts a reduction of -0.25%. Figure 10b for irrigated crops indicated that the oil is under-predicted by the emulator, while emulator predictions produce for cereal, rice and maize are relatively similar to LPJmL values. Uncertainty interval is much higher in oil than for other crops. Comparing the two plots irrigated rice and maize global percentage change in yield are higher than for rainfed

Table 2: Comparison of GP with WLS methods for four GCMs, with CO<sub>2</sub> fertilization, management level 5, RCPs 4.5 and 8.5, and all time slices for rainfed crops. The values are the proportion of variance  $\rho$  explained.

Crop	CCSR-MIROCMED		GISS-MODELER		GISS-MODELEH		IPSL-CM4	
	GP	WLS	GP	WLS	GP	WLS	GP	WLS
Cereal	0.87	0.81	0.71	0.66	0.78	0.72	0.75	0.72
Rice	0.83	0.81	0.71	0.67	0.83	0.67	0.71	0.72
Maize	0.92	0.86	0.69	0.67	0.76	0.68	0.87	0.80
Oil <sup>1</sup>	0.89	0.83	0.84	0.78	0.83	0.74	0.83	0.77

<sup>1</sup> Oil=yield<sub>max</sub>[soybean, rapeseed, sunflower].

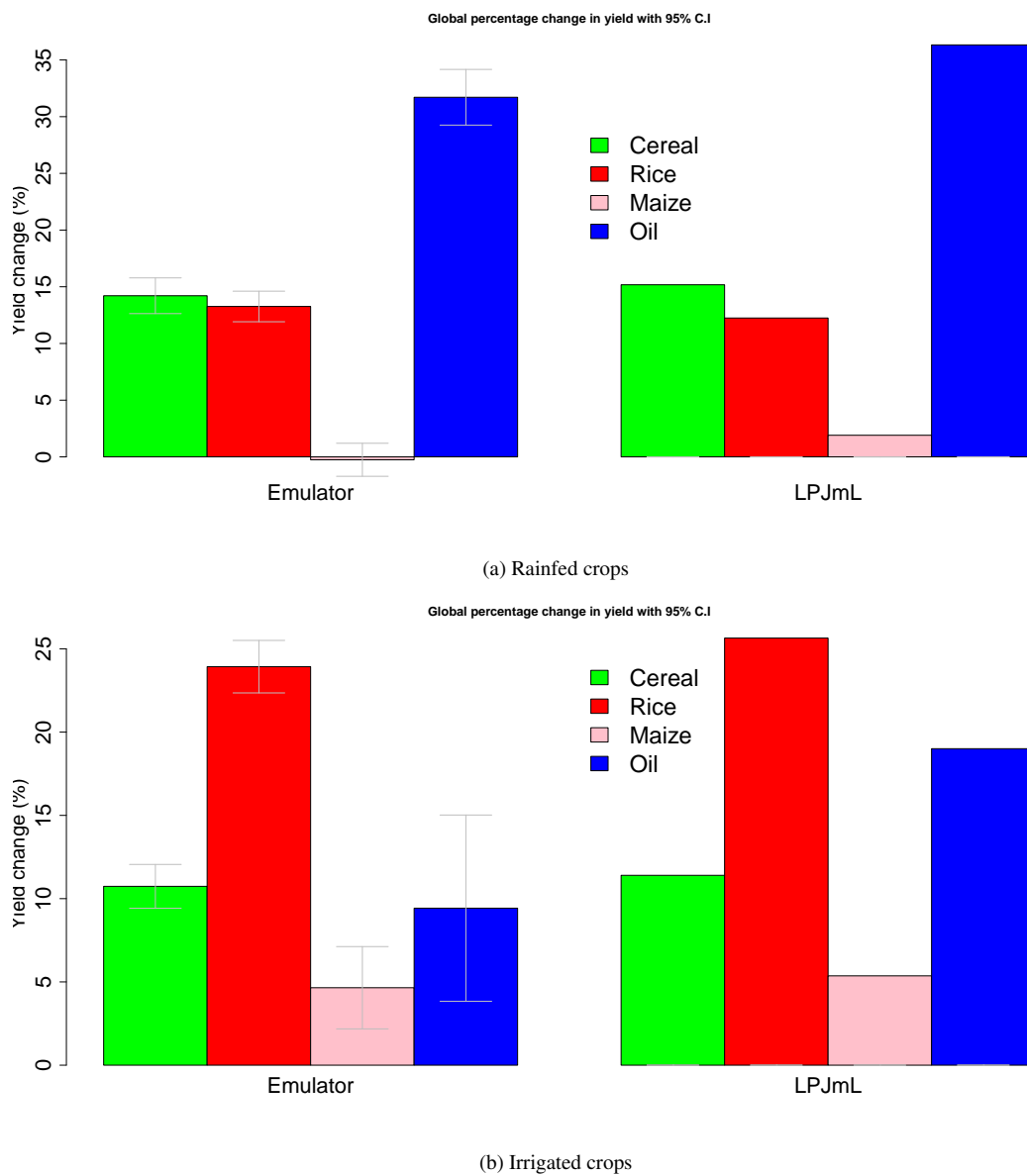


Fig. 10: Global percentage change in yield by emulator with its 95% C.I. in comparison with LPJmL values for a period (2085-2094) relative to (2005 – 2014) with their 95% C.I. The plots correspond to crop-yield for management level of 5, RCP6 with CO<sub>2</sub> fertilization from UKMO-HADGEM1.

rice and maize. In contrast, irrigated cereal and oil percentage changes are comparatively less than for rainfed crops.

## 5 Conclusion

In this paper, we have demonstrated a means of making inference about the parameters of the emulator using a GP regression that is based upon Bayesian principle. We have presented a simple statistical method for emulating the underlying physical dynamics of response of potential crop yield to change in climate. We use combinations of stepwise regression, PCA and GP methods to emulate major crop-yield as a linear function of seasonal climate variables and other relevant variables. Aggregated country average of actual yield change is calculated by combining yields simulated by LPJmL for both rainfed and irrigated. We consider eight time-slices averaged over ten years, baseline corresponding to the average (2005-2014).

In modelling the residual  $\mathbf{Z}$ , we reduced the complexity of the computation by aggregating the stepwise regression residuals from a fine to a more coarse resolution on a country level. We assume that the aggregation will reduce the complexity and structure of the global trend component of the emulator. We have already seen from the residual plots that some variations have not been accounted for by the stepwise regression model.

We demonstrated in [Oyebamiji et al. \(2015\)](#) that the stepwise regression approach captures some proportion of the global variation in our data. The residual analysis is performed to capture the remaining spatial local variation left unexplained in stage 1. In Bayesian framework, we assumed that our unexplained residual from stage1 is a GP. Under GP regression, we specified both the prior mean and covariance functions and estimated the hyperparameters of these functions by integrating out the nuisance parameters using the Bayes rule.

It is important to reduce the dimensionality of data for tractable application of GP modelling. We derived the input matrix  $\tilde{\mathbf{U}}$  from the stage 1 predictions, this matrix was then used as explanatory variables for GP regression. We first obtained the residual on a lower axis by aggregating to a country level. We fitted an independent GP in parallel to each column of  $\mathbf{Z}$ .

These emulators can predict the change in global crop yield as a function of climate from any GCM/RCP combination and for different management level options and CO<sub>2</sub> fertilization level. We provided some spatial plots to compare the performance of the emulators with the LPJmL.

Overall, cereal is more predictable in South America, Africa and China than other part of the world. Similarly, rainfed rice is more predictable in North America, China and India but is more difficult to predict in Australia. Maize and oil are generally much predictable in most regions except in Canada and Mongolia. Uncertainty levels are also relatively low for all the four crops. There are some isolated large confidence band in some countries which could be attributed to a large variability in the training data for these countries. The large uncertainty can be traced to inter-scenario variabilities across RCP and GCM.

Table 3: List of UN countries used for this analysis

Index	Country	Index	Country	Index	Country
1	Ocean	63	French Polynesia	125	New Zealand
2	Afghanistan	64	French Southern Territories	126	Nicaragua
3	Albania	65	Djibouti	127	Niger
4	Antarctica	66	Gabon	128	Nigeria
5	Algeria	67	Georgia	129	Norway
6	Angola	68	Gambia	130	Pakistan
7	Azerbaijan	69	State of Palestine	131	Panama
8	Argentina	70	Germany	132	Papua New Guinea
9	Australia	71	Ghana	133	Paraguay
10	Austria	72	Kiribati	134	Peru
11	Bahamas	73	Greece	135	Philippines
12	Bangladesh	74	Greenland	136	Poland
13	Armenia	75	Guatemala	137	Portugal
14	Belgium	76	Guinea	138	Guinea-Bissau
15	Bhutan	77	Guyana	139	Timor-Leste
16	Bolivia	78	Haiti	140	Puerto Rico
17	Bosnia and Herzegovina	79	Honduras	141	Qatar
18	Botswana	80	Hungary	142	Romania
19	Brazil	81	Iceland	143	Russian Federation
20	Belize	82	India	144	Rwanda
21	Solomon Islands	83	Indonesia	145	Saint Vincent and Grenadines
22	Brunei Darussalam	84	Iran	146	Saudi Arabia
23	Bulgaria	85	Iraq	147	Senegal
24	Myanmar	86	Ireland	148	Serbia
25	Burundi	87	Israel	149	Sierra Leone
26	Belarus	88	Italy	150	Slovakia
27	Cambodia	89	Cote d'Ivoire	151	Viet Nam
28	Cameroon	90	Jamaica	152	Slovenia
29	Canada	91	Japan	153	Somalia
30	Cape Verde	92	Kazakhstan	154	South Africa
31	Cayman Islands	93	Jordan	155	Zimbabwe
32	Central African Republic	94	Kenya	156	Spain
33	Sri Lanka	95	Democratic Republic of Korea	157	South Sudan
34	Chad	96	Republic of Korea	158	Sudan
35	Chile	97	Kuwait	159	Suriname
36	China	98	Kyrgyzstan	160	Swaziland
37	Taiwan	99	Lao Republic	161	Sweden
38	Colombia	100	Lebanon	162	Switzerland
39	Comoros	101	Lesotho	163	Syrian Arab Republic
40	Congo	102	Latvia	164	Tajikistan
41	D.R Congo	103	Liberia	165	Thailand
42	Costa Rica	104	Libya	166	Togo
43	Croatia	105	Lithuania	167	Trinidad and Tobago
44	Cuba	106	Luxembourg	168	United Arab Emirates
45	Cyprus	107	Madagascar	169	Tunisia
46	Czech Republic	108	Malawi	170	Turkey
47	Benin	109	Malaysia	171	Turkmenistan
48	Denmark	110	Mali	172	Uganda
49	Dominican Republic	111	Mauritania	173	Ukraine
50	Ecuador	112	Mauritius	174	The former Yugoslav
51	El Salvador	113	Mexico	175	Egypt
52	Equatorial Guinea	114	Mongolia	176	United Kingdom
53	Ethiopia	115	Republic of Moldova	177	United Republic of Tanzania
54	Eritrea	116	Montenegro	178	United States of America
55	Estonia	117	Morocco	179	United States Virgin Islands
56	Faeroe Islands	118	Mozambique	180	Burkina Faso
57	Falkland Islands (Malvinas)	119	Oman	181	Uruguay
58	South Georgia and South SS	120	Namibia	182	Uzbekistan
59	Fiji	121	Nepal	183	Venezuela
60	Finland	122	Netherlands	184	Samoa
61	Aland Islands	123	New Caledonia	185	Yemen
62	France	124	Vanuatu	186	Zambia

## Appendix 1: GP derivation

Let  $\mathbf{z} = [(z_i, g(u_i)) | i = 1, \dots, n]^T$  be some simulation data and given some prior distribution for hyperparameters. Then

$$P(\beta, \sigma^2 | \mathbf{z}) = \frac{P(\mathbf{z} | \beta, \sigma^2) P(\beta, \sigma^2)}{P(\mathbf{z})} \propto P(\mathbf{z} | \beta, \sigma^2) P(\beta, \sigma^2) \quad (14)$$

where  $P(\beta, \sigma^2 | \mathbf{z})$  is the parameter posterior distribution,  $P(\mathbf{z} | \beta, \sigma^2)$  is the likelihood,  $P(\beta, \sigma^2)$  the prior distribution for hyperparameters. The marginal likelihood or normalising factor  $P(\mathbf{z})$  can also be expressed as

$$P(\mathbf{z}) = \int P(\mathbf{z} | \beta, \sigma^2) P(\beta, \sigma^2) d\beta d\sigma^2 \quad (15)$$



under a full Bayesian framework, the posterior distribution in equation 14 and the marginal distribution in equation 15 are often difficult to compute for any arbitrary prior distribution, they may not exist in closed form. Conjugate or non-informative priors can be used to overcome this non-analyticity. A conjugate distribution arises when both the posterior and prior distributions come from the same family. Suppose, we want to make predictions at point  $u^{new}$  given some training data  $\mathbf{z}$ , then the joint distribution of the observed values  $\mathbf{z}$  and test point  $z^{new}$  under a given prior can be obtained.  $\mathbf{z}$  is

$$\mathbf{z} \sim \mathbf{N}(\mathbf{H}\beta, \sigma^2 \mathbf{C}) \quad (16)$$

and  $z^{new} = g(u^{new})$  have a joint multivariate Gaussian distribution using multivariate identity. Then we have

$$z^{new} | \mathbf{z}, \beta, \sigma^2 \sim \mathbf{N}[\mathbf{m}^*(.), \sigma^2 \mathbf{C}^*(.,.)]. \quad (17)$$

such that

$$\begin{aligned} m^*(\mathbf{u}) &= h^T(\mathbf{u})\beta + \mathbf{c}\mathbf{C}^{-1}[\mathbf{z} - \mathbf{H}(\mathbf{u})]\beta \\ C^*(\mathbf{u}, \mathbf{u}') &= c(\mathbf{u}, \mathbf{u}') - \mathbf{c}(\mathbf{u})^T \mathbf{C}^{-1} \mathbf{c}(\mathbf{u}'), \end{aligned}$$

In order to make predictions, we integrate over the posterior distribution (i.e. the expectation of the posterior density). Therefore, the posterior predictive density is given by

$$P(z^{new} | \mathbf{z}) = \int_{d\beta} \int_{d\sigma^2} P(z^{new} | \beta, \sigma^2) P(\beta, \sigma^2 | \mathbf{z}) d\beta d\sigma^2. \quad (18)$$

As mentioned, it can be difficult to obtain the  $P(z^{new} | \mathbf{z})$  analytically. The joint prior distribution  $P(\beta, \sigma^2 | \mathbf{z})$  in equation 18 can be further re-expressed below using the probability product identity

$$P(\beta, \sigma^2 | \mathbf{z}) = P(\beta | \sigma^2, \mathbf{z}) P(\sigma^2 | \mathbf{z}). \quad (19)$$

Let the prior distribution on the regression parameter  $\beta$  and  $\sigma^2$  be given by a weak prior such that

$$P(\beta, \sigma^2) \propto \frac{1}{\sigma^2} \quad (20)$$

Substituting equations (16 and 20) to the posterior distribution in equation 14 and with little algebraic manipulation we can infer that  $P(\beta | \sigma^2, \mathbf{z}) \sim N(\hat{\beta}, \sigma^2 (HC^{-1}H)^{-1})$  and  $\sigma^2 | \mathbf{z}$  has an inverse-gamma distribution where  $P(\sigma^2 | \mathbf{z}) = \int P(\sigma^2, \beta | \mathbf{z}) d\beta \sim (n - p - 2) \hat{\sigma}^2 \chi_{(n-p)}^{-2}$ . Let  $\hat{\beta}$  and  $\hat{\sigma}^2$  be the MLE values of  $\beta$  and  $\sigma^2$ , to be discussed shortly in section 5. Maximum likelihood estimate of the hyperparameters is an approximation to the true Bayesian method of integrating over all hyperparameters values weighted by their posterior. Therefore, equation 18 can be expressed as

$$\begin{aligned} P(z^{new} | \mathbf{z}) &= \int_{d\beta} \int_{d\sigma^2} P(z^{new} | \beta, \sigma^2) P(\beta | \sigma^2, \mathbf{z}) P(\sigma^2 | \mathbf{z}) d\beta d\sigma^2, \\ &\approx \int_{d\beta} \int_{d\sigma^2} P(z^{new} | \beta, \sigma^2) P(\beta | \sigma^2, \mathbf{z}) (n - p - 2) \hat{\sigma}^2 \chi_{(n-p)}^{-2} d\beta d\sigma^2 \\ &= (n - p - 2) \hat{\sigma}^2 \chi_{(n-p)}^{-2} \int_{d\beta} P(z^{new} | \beta, \hat{\sigma}^2) P(\beta | \hat{\sigma}^2, \mathbf{z}) d\beta. \end{aligned} \quad (21)$$

The values of the integrand can be evaluated noting that the first term in the integral follows from equation (17), and also recognise that  $P(\beta | \hat{\sigma}^2, \mathbf{z}) \sim N(\hat{\beta}, \hat{\sigma}^2 (HC^{-1}H)^{-1})$  (product of two Gaussian distribution). Therefore,

$$P(z^{new} | \mathbf{z}) = (n - p - 2) \hat{\sigma}^2 \chi_{(n-p)}^{-2} \times N[m^\bullet(.), \sigma^2 C^\bullet(.,.)]. \quad (22)$$

where

$$m^\bullet(\mathbf{u}) = h^T(\mathbf{u})\hat{\beta} + \mathbf{c}\mathbf{C}^{-1}[\mathbf{z} - \mathbf{H}(\mathbf{u})\hat{\beta}]$$

$$C^\bullet(\mathbf{u}, \mathbf{u}') = C^*(\mathbf{u}, \mathbf{u}') + \left[ (\mathbf{h}(\mathbf{u})^T \mathbf{c}(\mathbf{u}) \mathbf{C}^{-1} \mathbf{H}) (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} (\mathbf{h}(\mathbf{u})^T \mathbf{c}(\mathbf{u}) \mathbf{C}^{-1} \mathbf{H})^T \right],$$

$$P(z^{new} | \mathbf{z}) = T_{(n-p)} \left( m^\bullet(\mathbf{u}), \frac{(n-p) C^\bullet(\mathbf{u}, \mathbf{u}')}{(n-p-2)} \right). \quad (23)$$

where  $T_{(n-p)}$  is a student  $t$  distribution with  $(n-p)$  degree of freedom (Oakley, 1999, Murphy, 2007, Flores et al., 2014).

### Estimation of prior hyperparameters

Noting that under GP regression, the prior distribution for the data is also a Gaussian distribution. Let the log-likelihood of the parameters be given as

$$L(\beta, \sigma^2, \alpha) = -\frac{1}{2} \left[ n \log(\sigma^2) + \log(\det(\mathbf{C})) + (\mathbf{z} - \mathbf{H}\beta)^T \mathbf{C}^{-1} (\mathbf{z} - \mathbf{H}\beta) / \sigma^2 \right] \quad (24)$$

where  $\alpha = [\alpha_1, \dots, \alpha_n]$  is a vector of correlation lengths and  $\det(\mathbf{C})$  is the determinant of correlation matrix  $\mathbf{C}$ . Maximizing this function with respect to  $\beta$  will give the MLE of  $\beta$  as (also equal to its generalized least squares estimate)

$$\hat{\beta} = (\mathbf{H} \mathbf{C}^{-1} \mathbf{H})^T \mathbf{H} \mathbf{C}^{-1} \mathbf{z} \quad (25)$$

and for  $\sigma^2$  as

$$\hat{\sigma}^2 = \frac{1}{n} \left[ (\mathbf{z} - \mathbf{H}\hat{\beta})^T \mathbf{C}^{-1} (\mathbf{z} - \mathbf{H}\hat{\beta}) \right] \quad (26)$$

In order to compute the MLE of  $\alpha$ , the estimates of  $\hat{\beta}$  and  $\hat{\sigma}^2$  are substituted in equation (24) and maximized over  $\beta$  and  $\sigma^2$  to obtain

$$L(\hat{\beta}, \hat{\sigma}^2, \alpha) = -\frac{1}{2} \left[ n \log(\hat{\sigma}^2)(\alpha) + \log(\det(\mathbf{C}(\alpha))) + n \right], \quad (27)$$

which is a function of  $\alpha$  that can be computed using a derivative-free numerical optimisation like Nelder-Mead method see further details in (Santner et al., 2003, Rasmussen & Williams, 2006). Posterior distribution is obtained as  $P(\mathbf{z} | \hat{\beta}, \hat{\sigma}^2, \hat{\alpha}) \sim N[m^\bullet(\mathbf{u}), \sigma^2 C^\bullet(\mathbf{u}, \mathbf{u}')]$ , with posterior mean function given as

$$m^\bullet(\mathbf{u}) = h^T(\mathbf{u}) \hat{\beta} + \mathbf{c}(\mathbf{u}) \mathbf{C}^{-1} [\mathbf{z} - \mathbf{H}(\mathbf{u}) \hat{\beta}] \quad (28)$$

and posterior covariance

$$C^\bullet(\mathbf{u}, \mathbf{u}') = C^*(\mathbf{u}, \mathbf{u}') + \left[ (\mathbf{h}(\mathbf{u})^T \mathbf{c}(\mathbf{u}) \mathbf{C}^{-1} \mathbf{H}) (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} (\mathbf{h}(\mathbf{u})^T \mathbf{c}(\mathbf{u}) \mathbf{C}^{-1} \mathbf{H})^T \right], \quad (29)$$

Advantage of GP method is that the prediction is probabilistic by providing confidence interval as a measure of uncertainty. GP is a robust method because it allows for different correlation and linear regression functions. Additional information in the form of a prior distribution improves the prediction. However, its efficiency reduces for high-dimensional data (scale poorly).

### Appendix 2: Model performance

We compute the squared differences between the actual LPJmL values  $y$  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

$$\rho = 1 - \frac{\sum_{i=1}^4 \sum_{j=1}^2 \sum_{k=1}^7 \sum_{t=1}^8 \sum_{n=1}^N (y_{ijktn} - \tilde{y}_{ijktn}^*)^2}{\sum_{i=1}^4 \sum_{j=1}^2 \sum_{k=1}^7 \sum_{t=1}^8 \sum_{n=1}^N (y_{ijktn} - \bar{y})^2} \quad (30)$$

and the overall cross-validation root mean squared error (RMSE<sub>CV</sub>) is

$$\text{RMSE}_{CV} = \left( \sum_{i=1}^4 \sum_{j=1}^2 \sum_{k=1}^7 \sum_{t=1}^8 \sum_{n=1}^N \frac{(y_{ijktn} - \tilde{y}_{ijktn}^*)^2}{(4 \times 2 \times 7 \times 8 \times N)} \right)^{1/2}. \quad (31)$$

Table 4: The emulator's input variables for stepwise regression.

Variables	Full names
	2001-2100 data
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	Change in mean wet day frequency in summer/winter/spring/autumn
iscld/iweld/ispclld/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
CO <sub>2</sub> and CCO <sub>2</sub>	Initial (baseline) mean CO <sub>2</sub> and change in mean CO <sub>2</sub>
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.

## References

- Arnell, N. W., Lowe, J. A., Brown, S., Gosling, S. N., Gottschalk, P., Hinkel, J., ... & Warren, R. F. (2013). A global assessment of the effects of climate policy on the impacts of climate change. *Nature Climate Change*, 3(5), 512 – 519. [2](#)
- Bates, B.C., Kundzewicz, Z.W., Wu, S., Palutikof, J.P. (Eds.) (2008). Climate Change and Water. *IPCC Technical Paper VI*, IPCC, Geneva. 210 pp. [13](#)
- Bayarri, M., Berger, J., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R., Paulo, R., Sacks, J. and Walsh, D. (2007). Computer model validation with functional output. *Annals of Statistics*, 35(5), 1874 – 1906.
- Bondeau, A., Smith, P.C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Miller, C., Reichstein, M., Smith, B. (2007). Modelling the role of agriculture for the 20th century global terrestrial carbon balance, *Global Change Biology*, 13(3), 679 – 706. [3](#)
- Bornn, L., Zidek, J.V. (2012). Efficient stabilization of crop yield prediction in the Canadian Prairies. *Agricultural & Forest Meteorology*, 152, 223 – 232. [5](#)
- Boukouvalas, A., Cornfold, D. (2008). Dimension reduction for multivariate emulation. *Technical report*, NCRG/006. [3](#)
- Conti, S., Gosling, J. P., Oakley, J. E., & O'Hagan, A. (2009). Gaussian process emulation of dynamic computer codes. *Biometrika*, 96(3), 663 – 676.
- Conti, S., & OHagan, A. (2010). Bayesian emulation of complex multi-output and dynamic computer models. *Journal of statistical planning and inference*, 140(3), 640 – 651.
- Dancik, G. (2011). mlegp: an R package for Gaussian process modeling and sensitivity analysis. [5](#)
- David G.T. Denison (Ed.). (2002). Bayesian methods for nonlinear classification and regression (Vol. 386). John Wiley & Sons.
- Drineas, P. and Mahoney, M.W. (2005). On the Nystrom Method for approximating a Gram matrix for improved kernel-based learning. *Journal of Machine Learning Research*, 6, 2153 – 2175. [5](#)
- Finley, A. O., Banerjee, S., & Basso, B. (2011). Improving crop model inference through Bayesian melding with spatially varying parameters. *Journal of agricultural, biological, and environmental statistics*, 16(4), 453 – 474. [5](#)
- Floresa, E.I., DiazDelaO, F.A., Friswell, M.I., Ajaj, R.M. (2014). Investigation on the extensibility of the wood cell-wall composite by an approach based on homogenisation and uncertainty analysis. *Composite Structures*, 212 – 222. [22](#)
- Fricker, T., Oakley, J., & Urban, N. M. (2010). Multivariate emulators with nonseparable covariance structures.
- Gosling, J.P. and O'Hagan, A. (2007). Understanding the uncertainty in the biospheric carbon flux for England and Wales. *Technical Report*, 567/06, Department of Probability and Statistics, University of Sheffield. [2](#)
- Hankin, R. K. (2012). Introducing multivator: A Multivariate Emulator. *Journal of Statistical Software*, 46(8), 1 – 20.
- Heitmann, K., Higdon, D., Nakhleh, C., Habib, S., (2006). Cosmic Calibration. *The Astrophysical Journal*, 646, 2, L1 – L4. [3](#), [5](#)
- Higdon, D., Gattiker, J., Williams, B. & Rightley, M. (2008). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103, 570 – 583. [2](#), [3](#), [5](#)
- IPCC WGI Fourth Assessment Report (2007): Climate Change: *The Physical Science Basis*, in Summary for Policymakers by Alley, R., Berntsen, T., Nathaniel, R., Bindoff, L. et al. [1](#)
- IPCC, 2014: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken,

- P.R. Mastrandrea, and L.L. White (eds.)). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1132 pp. [2](#)
- Jain, R.C. and Agrawal, R. (1992). Probability model for crop yield forecasting. *Biometrical Journal*, 34(4), 501 – 511. [2](#)
- Kennedy, M.C. & O’Hagan, A. (2001). Bayesian calibration of computer models. *Journal of Royal Statistical Society*, series B, 63(3), 425 – 464. [3](#)
- Kennedy, M. C., Anderson, C. W., Conti, S., and O’Hagan, A. (2006). Case studies in Gaussian process modelling of computer codes. *Reliability Engineering & System Safety*, 91, 1301 – 1309. [3](#)
- Kennedy, M.C. et al. (2008). Quantifying uncertainty in the biospheric carbon flux for England and Wales. *Journal of the Royal Statistical Society*, Series B, 171, 109 – 135. [2](#)
- Kim, Y., Lee, J., Kim, J. (2005). Bayesian bootstrap analysis of doubly censored data using Gibbs sampler. *Statistica Sinica*, 969 – 980. [2](#)
- Kitchen, N.R., Sudduth, K.A., Drummond, S.T. (1999.) Electrical conductivity as a crop productivity measure for claypan soils. *Journal of Prod. Agriculture*, 12(4), 607617. [2](#)
- Lee, L. A., Carslaw, K. S., Pringle, K. J., & Mann, G. W. (2012). Mapping the uncertainty in global CCN using emulation. *Atmospheric Chemistry and Physics*, 12(20), 9739 – 9751. [5](#)
- Lee, L. A., Pringle, K. J., Reddington, C. L., Mann, G. W., Stier, P., Spracklen, D. V., & Carslaw, K. S. (2013). The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei. *Atmospheric Chemistry & Physics Discussions*, 13(3). [5](#)
- Lopez, A., Tebaldi, C., New, M., Stainforth, D., Allen, M., Kettleborough, J. (2006). Two approaches to quantifying uncertainty in global temperature changes. *Journal of Climate*, 19, 4785 – 4796. [2](#)
- Matis, J.H., Birkett, T. and Boudreaux, D. (1989). An application of the Markov chain approach to forecasting cotton yields from Surveys. *Agricultural Systems*, 29, 357 – 370. [2](#)
- Meinshausen, M., Raper, S.C., and Wigley, T.M. (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.
- Morris, J. S., Vannucci, M., Brown, P. J. & Carroll, R. J. (2003). Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis. *Journal of the American Statistical Association*, 98, 573 – 583.
- Morris, J. S., & Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2), 179 – 199.
- Murphy, K. P. (2007). Conjugate Bayesian analysis of the Gaussian distribution. *Technical Report*, University of British Columbia. [22](#)
- Oakley, J. (1999). Bayesian Uncertainty Analysis For Complex Computer Codes. Ph.D. thesis, University of Sheffield. [22](#)
- Oakley, J. and O’Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89, 769 – 784. [2](#)
- Oakley, J. E. and O’Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of Royal Statistical Society*, 66B, 751 – 769. [3](#)
- Oyebamiji, O.K., Edwards, N.R., Holden, P.B., Garthwaite, P.B., Schaphoff, S., and Gerten, D. (2015). Emulating global climate change impacts on crop yields. *Statistical Modelling*, 1471082X14568248. [2](#), [3](#), [6](#), [19](#)
- 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. *Agricultural & Forest Meteorology*, 170, 183 – 194. [2](#)
- Parry, M., Rosenzweig, C., & Livermore, M. (2005). Climate change, global food supply and risk of hunger. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1463), 2125 – 2138. [13](#)
- Parry, M. L. (Ed.). (2007). Climate Change 2007: impacts, adaptation and vulnerability: contribution of Working Group II to the fourth assessment report of the Intergovernmental Panel on Climate Change (Vol. 4).

- Cambridge University Press. [13](#)
- Quinonero-Candela, J., & Rasmussen, C. E. (2005). A unifying view of sparse approximate Gaussian process regression. *The Journal of Machine Learning Research*, 6, 1939 – 1959. [4](#)
- 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/>.
- Rasmussen, C.E. and Williams, C.K.I. (2006). Gaussian Processes for Machine Learning, the MIT Press. [4](#), [5](#), [22](#)
- Rosenzweig C, Parry M.L. (1994). Potential impact of climate change on world food supply. *Nature*, 367, 133 – 138. [13](#)
- Rougier, J.C. (2007). Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climate Change*, 81, 247 – 264. [2](#)
- Rougier, J. (2008). Efficient emulators for multivariate deterministic functions. *Journal of Computational and Graphical Statistics*, 17(4), 827 – 843. [5](#)
- Rougier, J., Guillas, S., Maute, A., & Richmond, A. D. (2009). Expert knowledge and multivariate emulation: The thermosphere-ionosphere electrodynamics general circulation model (TIE-GCM). *Technometrics*, 51(4), 414 – 424. [5](#)
- Sacks, J., Welch, W., Mitchell, T., Wynn, H. (1998). Design and analysis of computer experiments. *Statistical Science*, 4(4), 409 – 435. [4](#)
- Sacks, W.J., Deryng, D., Foley, J.A., Ramankutty, N. (2010). Crop planting dates: an analysis of global patterns. *Global Ecology and Biogeography*, 19, 607-620. [2](#)
- Santner, T., Williams, B., Notz, W. (2003). The Design and Analysis of Computer Experiments. Springer. [4](#), [22](#)
- Schlenker, W., Michael J. Roberts, M.J. (2006). Estimating the impact of climate change on crop yields: The importance of non-linear temperature effects. *Discussion Paper*, 0607-01, Department of Economics Columbia University New York, NY. [2](#)
- Shi, J.Q., Murray-Smith, R. and Titterton, D.M. (2005). Hierarchical Gaussian Process mixtures for regression. *Statistics and Computing*, 15, 31 – 41. [4](#)
- Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). *Journal of Royal Statistical Society, series B*, 47(1), 1 – 52. [5](#)
- Sitch, S., Smith, B., Prentice I.C., 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. [3](#)
- Solomon, S. (2007) (eds) Climate Change: The Physical Science Basis. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, UK.
- Stocker, T. F., Dahe, Q., & Plattner, G. K. (2013). Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers (IPCC, 2013). [13](#)
- Tebaldi, C., Smith, R., Nychka, D., Mearns, L. (2005). Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles. *Journal of Climate*, 18, 1524 – 1540. [2](#)
- Wahba, G. (1990). Spline Models for Observational Data. Society for Industrial and Applied Mathematics, Philadelphia. [5](#)
- Wahba, G., Johnson, D. R., Gao, F., and Gong, J. (1995). Adaptive tuning of numerical weather prediction models: Randomized GCV in three-and four-dimensional data assimilation. *Monthly Weather Review*, 123, 3358 – 3369. [5](#)
- Wallach D. (2011). Crop model calibration: A statistical perspective. *Agronomy Journal*, 103(4), 1144 – 1151. [2](#)



- Warren, R., Lowe, J. A., Arnell, N. W., Hope, C., Berry, P., Brown, S., ... & Vanderwal, J. (2013). The AVOID programmes new simulations of the global benefits of stringent climate change mitigation. *Climatic Change*, 120(1 – 2), 55 – 70. [2](#)
- Wilkinson, R.D. (2010). Bayesian calibration of expensive multivariate computer experiments. Large-Scale Inverse Problems and Quantification of Uncertainty, *Ser. Comput. Stat.*, edited by LT Biegler et al, 195 – 216.
- Wilkinson, R.D., in: Biegler et al. (Eds.) (2010). Large-Scale Inverse Problems and Quantification of Uncertainty. *John Wiley & Sons*, New York. [3](#)
- Wu, W., Verburg, P. H., & Tang, H. (2014). Climate change and the food production system: impacts and adaptation in China. *Regional Environmental Change*, 14(1), 1-5. [13](#)
- Young, P.C. and Ratto, M. (2011). Statistical emulation of large linear dynamic models. *Technometrics*, 53(1), 29 – 43. [3](#)
- Zhou, Y., Li, N., Dong, G., & Wu, W. (2013). Impact assessment of recent climate change on rice yields in the Heilongjiang Reclamation Area of north-east China. *Journal of the Science of Food and Agriculture*, 93(11), 2698 – 2706. [13](#)