CLIMATE CHANGE AND AGRICULTURAL PRODUCTIVITY IN SUB-SAHARAN AFRICA: A SPATIAL SAMPLE SELECTION MODEL

by

Patrick S. Ward, Raymond J.G.M. Florax, and Alfonso Flores-Lagunes

Working Paper # 11-4
September 2011

Dept. of Agricultural Economics
Purdue University

CLIMATE CHANGE AND AGRICULTURAL PRODUCTIVITY IN SUB-SAHARAN AFRICA: A SPATIAL SAMPLE SELECTION MODEL

by

Patrick S. Ward[1], Raymond J.G.M. Florax[1,2,3], and Alfonso Flores-Lagunes[4,5]

[1] Department of Agricultural Economics, Purdue University
West Lafayette, IN 47907-2056 USA

[2] Department of Spatial Economics, VU University
Amsterdam, The Netherlands

[3] Tinbergen Institute
Amsterdam, The Netherlands

[4] Department of Economics, Binghamton University, State University of New York
Binghamtom, NY 13902-6000 USA

[5] Institute for the Study of Labor (IZA)
Bonn, Germany

pward@purdue.edu, rflorax@purdue.edu, aflores@binghamton.edu

Working Paper 11-4

September 2011

Abstract

Using data at a high spatial resolution, we estimate a cereal yield response function conditional upon climatological and topographical features using a recently developed estimator for spatial process models when sample selection is of concern. We control for localized spatial correlation in unobserved disturbances affecting both the selection to plant cereals as well as in the resulting conditional yield response. We find that cereal yields across Sub-Saharan Africa will decline with increasing temperatures resulting from global climate change, and that failing to control for sample selection leads to underestimation of these adverse effects.

Keywords: Agricultural productivity, climate change, spatial econometrics, sample selection, generalized method of moments

JEL Codes: C31, Q18, C50

Climate Change and Agricultural Productivity in Sub-Saharan Africa: A Spatial Sample Selection Model*

Patrick S. Ward
Department of Agricultural Economics, Purdue University, USA

Raymond J.G.M. Florax
Department of Agricultural Economics, Purdue University, USA
Department of Spatial Economics, VU University Amsterdam, The Netherlands
and Tinbergen Institute, The Netherlands

Alfonso Flores-Lagunes
Department of Economics, Binghamton University, SUNY, USA
and IZA, Bonn, Germany

September 29, 2011

Abstract

Using data at a high spatial resolution, we estimate a cereal yield response function conditional upon climatological and topographical features using a recently developed estimator for spatial process models when sample selection is of concern. We control for localized spatial correlation in unobserved disturbances affecting both the selection to plant cereals as well as in the resulting conditional yield response. We find that cereal yields across Sub-Saharan Africa will decline with increasing temperatures resulting from global climate change, and that failing to control for sample selection leads to underestimation of these adverse effects.

JEL codes: C31, Q18, C50
Keywords: Agricultural productivity, climate change, spatial econometrics, sample selection, generalized method of moments

*Corresponding author: pward@purdue.edu. Special appreciation is extended to participants in sessions at the 3^{rd} Annual Midwest Regional, Population, and Health Graduate Student Summit, the IV^{th} World Conference of the Spatial Econometrics Association, the 2010 Agricultural and Applied Economics Association annual meeting, Gerald Shively, Jason Brown and Benoît Delbecq. This research was made possible, in part, through support provided by the Bureau of Economic Growth, Agriculture and Trade, U.S. Agency for International Development through the Global Nutrition and BASIS Assets and Market Access Collaborative Research Support Programs. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the sponsoring agency.

1 Introduction

In this paper we consider how climate change is expected to impact agricultural productivity in Sub-Saharan Africa, specifically with respect to cereal productivity. Cereals are an important part of a fundamental sector in a region of the world expected to be hard hit by climate change. Because the agricultural sector is inherently dependent upon climate conditions, and is often considered to be more susceptible to the adverse effects of global climate change than any other sector (e.g., WDR, 2010), there have been many previous studies that estimate the effects of anticipated climate changes on the agricultural sector.[1] Many of these previous studies are potentially flawed by ecological or sample selection bias. We control for these biases using an econometric method well-suited for analyzing high resolution, spatially-explicit samples that potentially exhibit non-randomness. Ultimately, we simulate productivity changes resulting from spatially heterogeneous changes in temperature and precipitation and test the effects on agricultural productivity of varying improvements to irrigation infrastructure.

To date, studies exploring the effects of climate change on the agricultural sector have typically fallen into one of three general categories: agronomic crop models (Jones and Kiniry, 1986; Jones et al., 1991), computable general equilibrium (CGE) studies (Darwin et al., 1995), and statistical models (Mendelsohn et al., 1994; Mendelsohn and Dinar, 1999; Schlenker and Roberts, 2006; Schlenker et al., 2006; Lobell et al., 2008). Each of these three general approaches is unique and has its own strengths and weaknesses, and the appropriateness of a particular methodology in a particular context may be evaluated by the following criteria (Rowhani and Ramankutty, 2009): data requirements, spatial extent, spatial resolution, user friendliness, and process understanding.[2]

One of the most widely used empirical tools for estimating the impacts of climate change on agricultural output is the agronomic crop growth model. Agronomic crop growth models incorporate a mechanistic growth process in complex computer simulation models. Several commonly cited agronomic crop models include the Crop Environment Resource Synthesis (CERES) model, the Erosion Productivity Impact Calculator (EPIC) model, and the CROPGRO models. The Decision Support

[1]For a synopsis of the findings of this vast literature, the interested reader is directed to Cline (2007), Chapter 2.

[2]The latter refers to models that take into consideration the bio-physical processes by which plants grow.

System for Agrotechnology Transfer Cropping System Model (DSSAT-CSM) attempts to incorporate crop growth models with knowledge about the contributions of soil, climate, and agricultural management practices into a unified tool for making better crop management decisions. At its core, DSSAT is a collection of independent crop growth models, but it also incorporates databases describing weather, soil, genetics, pests, experimental conditions, and economic information useful in applying the growth models to various simulations (Jones et al., 2003). A noted strength of these models is that, because of their process-based nature, crop growth is simulated in stages, and the timing of climatological variations can be modeled in a manner more structurally consistent with true agronomic processes (Hertel and Rosch, 2010). A criticism of these analyses, particularly amongst economists, is that they do not incorporate adaptation to changing climate conditions. These models implicitly assume that farmers do not switch their cropping patterns in the face of changing climate conditions and yield responses, so economic impacts are generally predicated on the cultivation of the given crop. Consequently, crop growth models tend to overstate the damages of climate change to agricultural production, particularly in developing countries (Mendelsohn and Dinar, 1999). Agronomic crop models are generally the most data intensive, since they require detailed data on local crop management and climate. Because agronomic crop studies incorporate spatially-explicit climate conditions in controlled settings, resulting yield responses are spatially explicit. While such detailed, local level responses may be advantageous, the incorporation of observed crop responses into global models can be very complex.

Computable General Equilibrium (CGE) models allow for a greater deal of adaptation, since they allow for endogenous price movements to stimulate varying allocations in scarce resources. Particularly, the sub-class of Agro-Ecological Zone (AEZ) studies combines crop simulation models with land-use decision analysis, and model changes in agronomic resources to assess changes in agricultural production, premised on lands shifting from one agro-ecological classification to another under changes in environmental conditions (Cline, 2007). An early CGE model applied to analyzing the impacts of climate change on the agricultural sector is Darwin et al. (1995), who utilized a multi-region, multi-sector CGE framework to assess changes in land use from effects of climate change derived from four global circulation models. Because they allow for endogenous input and output prices, CGE studies explicitly recognize that land values may change as a result of climate change, and therefore the uses of land may change as well. These models are often calibrated or parameterized on the basis of statistical relationships and may be structured to take into consideration finite (and potentially immobile) resources, as well as intersectoral and international linkages. A drawback of these models is that they are often subject to rather severe aggregations, such that a great deal of spatial and economic heterogeneity is lost as individual actors are aggregated up until they are represented by "representative agents", who may bear little or no resemblance to the individual actors at the lowest level. Additionally, these models are not process-based, and results are inherently dependent upon behavioral assumptions and the subsequent parameterization of the model, as well as on assumptions regarding market clearance, rather than actual biophysical processes. Largely because of assumptions of fluid prices and market closure, Darwin et al. (1995) find that the adverse effect of climate change on yields will drive up food prices, ultimately resulting in increased land being devoted to food crop production, with little change in overall actual output. CGE models can be very difficult and time-consuming to construct, due to the exhaustive accounting and the complexity of modeling the intersectoral and international linkages, and these models often require a great deal of macroeconomic data, data on resource endowments, and bilateral trade flow data.

Statistical models provide somewhat of a middle ground. These models rely on the estimation of statistical relationships between environmental variables and various outcome variables. Within the literature examining the economic consequences of climate change on the agricultural sector, statistical models are typically organized in two strands: production functions and Ricardian models. The production function approach typically uses either cross-sectional or time-series data to express yields as a function of inputs, including environmental factors such as temperature and precipitation, labor and fertilizer inputs. A recent study by Lobell and Burke (2010) finds that studies based on time-series data better predict the responses of yields to precipitation changes than to temperature changes, while studies based on cross-sectional or panel data perform better at predicting the responses to temperature changes than to precipitation changes. Ricardian models are essentially cross-sectional, reduced-form hedonic models, using statistical methods to estimate the response of land values to climatic changes. The pioneering entry in this literature was Mendelsohn et al. (1994), who found that, in the United States higher temperatures generally resulted in lower land values. They therefore suggested that the effects of global warming on the agricultural sector might be lower than estimated. Because land rents are assumed to reflect the value of the activity to which that land is allocated, these models are thought to embody adaptation, thereby controlling for the "naive" farmer scenario that is identified in traditional production function and agronomic crop growth approaches. Implicitly, these models assume that farmers simply produce along the outer envelope of a hedonic value surface, where the value functions for different activities are represented as a function of temperature (or some other environmental variable). While this may embody some forms of adaptation, it does not, as is often suggested, embody all forms of adaptation. It only captures the forms of adaptation or substitution of land uses that were actually reflected in the data, with temporal variations assumed to be proxied by cross-sectional variations. An important criticism raised by Cline (1996) is that, because the only adaptations incorporated in the model are those actually observed in the data, these models implicitly assume an "infinitely elastic supply of irrigation water" at prevailing prices at the time of the study. Additionally, Cline (1996) remarks that, while these models suggest that land may be allocated towards its most profitable uses, such that land use shifts from wheat and corn into grazing and retirement homes, there is no mechanism to capture the fact that, if wheat and corn production are eventually replaced by these other uses, there will be insufficient food production to feed the residents of all of these retirement homes. Despite these apparent weaknesses, the various statistical models remain very popular, especially among economists.

Unlike agronomic crop growth models and CGE models, statistical models are generally simple to implement and have minimal data requirements. This is particularly true for statistical studies estimating production functions, since the most parsimonious model requires only historical information on crop yields and climate conditions. But while statistical models are generally the least data-intensive of these methods, they are often limited in their spatial resolution, using aggregated data on average yields or average climate conditions for administrative units. As was the case with CGE analyses, such aggregations potentially cause ecological bias, since such interpretations imply that relationships that hold for grouped data hold for lower-level units (Haining, 2003). In addition, yield responses for a country (or some other administrative unit) are based on observed yields within that administrative unit, assuming that the observed yields are randomly drawn from a population of yields within that administrative unit. If the observed yields are actually drawn from a different unobserved distribution, there is the potential for sample selection bias, and the inferences drawn from these observations should not be extrapolated to the entire administrative unit.

We attempt to correct for ecological and sample selection biases by incorporating high resolution, spatially-explicit data and modeling spatial autocorrelation, while at the same time controlling for potential sample selection bias. Our methodological approach is somewhat of a hybrid, blending the spatial nature of agronomic crop yield studies with the minimal data requirements and the simple cross-sectional econometric approach of traditional statistical production function models. Whereas the agronomic studies use process-based responses to simulate climate change impacts, our study uses historically-observed climate conditions and crop yields, and utilizes statistical methods to isolate *ceteris paribus* effects of explanatory factors in explaining yield responses. While we use data at the level of geographic $1^\circ \times 1^\circ$ grid cells to partially avoid ecological biases that would likely be present if our focus was at the country level, we acknowledge that there are unobservable factors at the country level that may be relevant in explaining differences in cereal yields across space. For this reason, we attempt to control for this high-level heterogeneity by including binary variables to capture these country-specific effects.

Next to South Asia, Sub-Saharan Africa is the most agriculture-dependent region in the world, with nearly 19 percent of the region's gross output being produced in the agricultural sector (World Bank, 2006). This figure understates the importance of agriculture across the region, since the largest economy in the region, South Africa, has a relatively low dependence on agriculture. Many of the countries in Sub-Saharan Africa derive upwards of 30 percent of their GDP from agricultural production. Particularly, cereals (wheat, rice, barley, maize, rye, oats, millet and sorghum) play an important role in the diets of people in Sub-Saharan Africa. Cereals constitute 47 percent of total caloric food consumption (Kcal/capita/day) for households in Sub-Saharan Africa and 50 percent of protein consumption FAO (2011). Additionally, cereals provide calories more cheaply than other sources of food (Deaton, 1997). Despite this heavy reliance on agriculture, and despite the continued importance of cereal grains, cereal yield growth in Sub-Saharan Africa has consistently lagged behind other regions (FAO, 2011).

There is significant heterogeneity in agricultural productivity growth among the various sub-regions of Sub-Saharan Africa. For example, yields in Southern Africa have shown the most growth since the early 1960s, with yields increasing more than 350 percent during the roughly 50 year period. This growth has, however, been very uneven and variable over time. Yield growth in other parts of Sub-Saharan Africa has been much slower, but also generally more steady. Yield growth in Middle Africa was relatively flat until the mid-1990s, and even since then yield growth has been inconsistent. As of 2009, average yields in Middle Africa were only 878 kg/ha, compared with over 3,700 kg/ha in Southern Africa (FAO, 2011). Some of these vast differences can be attributed to differences in climate and soil quality, but it is also highly probable that there are other factors that contribute to these differences. Some factors, such as production technology and factor inputs, would be expected to demonstrate moderately high degrees of spatial dependence. Other factors may be country-specific. Arguably, it is not a coincidence that yields and yield growth are lowest in Middle Africa, which contains several countries that have experienced prolonged armed conflicts, and which rank near the bottom in the World Bank's World Governance Indicators' ranking for political stability (Kauffmann et al., 2010).

The remainder of this paper is structured as follows. In section 2 we introduce the general model of interest and impose specific restrictions on this general model to arrive at our ultimate estimable relationship. In section 3, we introduce the data to be used in this analysis. In section 4, we introduce our estimation strategy and present our results, including an interpretation of the marginal effects of the various conditioning factors on yields. Using parameter estimates from our cereal yield response function in conjunction with spatially heterogeneous climate change projections (drawn from Cline, 2007), we conduct hypothetical policy experiments in section 5 to examine how changes in irrigation infrastructure would affect cereal yields. Finally, we offer concluding remarks in section 6.

2 Model Specification

Consider the following general cross-sectional spatial error model:

$$
\begin{aligned}
y_i &= \mathbf{x}_i'\beta + \varepsilon_i \\
\varepsilon_i &= \rho \sum_j \mathrm{w}_{ij}\varepsilon_j + u_i
\end{aligned}
\tag{1}
$$

where y_i is the dependent variable, $\mathbf{x}_i'$ is a k-vector of explanatory variables conditioning y_i, ε_i is an unobserved disturbance term, u_i is a normally distributed random error, w_{ij} is the (i, j) element of a spatial weights matrix that defines the structure of the spatial system, and ρ is a spatial correlation coefficient corresponding to the spatially autoregressive errors. We assume that the spatial autoregressive parameter and the spatial weights matrix satisfy the conditions common to most models of this form (see, for example, Kelejian and Prucha, 1998, 2010). These assumptions are required to ensure a well-behaved spatial processes and estimability. We maintain the assumption that the underlying spatial process is a spatial autoregressive error process.[3] While it is true that cereal yields are likely to be correlated across space, the spatial process among yields is not causal or indicative of spatial dependence. The spatial correlation among yields, therefore, is mostly due to correlation among unobserved factors (e.g., factors of production, production technology, or knowledge spillovers), which are captured more appropriately through a spatial error process. Furthermore, there is precedence in modeling crop yield response functions and Ricardian models as spatial error processes (Schlenker et al., 2006).

Because of the differences in yields and productivity growth across Sub-Saharan Africa, it is also of interest to consider spatial heterogeneity. In essence, spatial heterogeneity points to instability or non-stationarity in the spatial system. Anselin (1988) identified two primary sources of spatial heterogeneity: heteroskedastic errors and spatially varying parameters. While spatial models artificially introduce heteroskedasticity, it is generally preferable to also explicitly incorporate spatial heterogeneity in the model. Since the data, once aggregated, are at $1^\circ \times 1^\circ$ spatial resolution, it is logical to identify spatial shift operators based on the country to which each grid cell belongs.

[3] Anselin et al. (1996) developed a series Lagrange Multiplier tests to identify the underlying spatial process. These tests, however, are not valid when the data fail to satisfy strict distributional assumptions.

There are many factors (such as governmental subsidies for fertilizers, extension programs to assist farmers' decision-making, a sound legal system which provides enforceable property rights, a system of functioning input and output markets, etc.) that would plausibly be the same for all grid cells within a country, but would vary across countries. We therefore include country-specific fixed effects in the model.

It is also relevant to allow for parametric heterogeneity through the identification of spatial regimes. In a critique of earlier Ricardian models, Schlenker et al. (2005) suggest that the effects of climate change on agriculture must be assessed differently in dryland and irrigated areas. Failure to account for irrigation, they argue, understates the water supply in irrigated regions. They suggest that irrigation be included in the set of explanatory variables, but also that the effects of climatological variables should be different in irrigated areas than in rain-fed areas.

With these additional model specifications imposed, we can re-write equation (1) as:

$$\begin{aligned} y_i &= \mathbf{x}_i'\beta + \mathbf{h}_i'\pi + (\mathrm{Irr}_i \cdot \tilde{\mathbf{x}}_i)'\,\xi + \varepsilon_i \\ \varepsilon_i &= \rho \sum_j \mathrm{w}_{ij}\varepsilon_j + u_i \end{aligned} \tag{2}$$

where y_i is cereal yield per grid cell i (in tons per hectare), $\mathbf{x}_i'$ is a vector of explanatory variables for grid cell i containing an intercept term and observations on average temperature (°C) and its square, the standard deviation of temperature, average precipitation (mm per month) and its square, the standard deviation of precipitation, average elevation (m), the roughness of the terrain (a measure of variation in elevation), the distance to the shore (km), the percentage of the cell that has irrigation, the average pH level for the soil in the cell, and the average carbon content in the cell's soil. The vector $\mathbf{h}_i'$ is a vector of country dummy variables and vector $\tilde{\mathbf{x}}_i \subset \mathbf{x}_i$ is a vector of explanatory variables that are interacted with grid cell-level irrigation.[4] The vector π is a vector of parameters to be estimated capturing country-specific effects, and ξ is a vector of parameters capturing effects of interactions between climatological and soil characteristics and irrigation. This

[4]This vector excludes the quadratic temperature and precipitation terms, as well as elevation, roughness of terrain, distance to the shore, and the cell-level irrigation proportion. These interaction terms were not significant in explaining yields in a simple OLS regression, and it seems intuitive that the effect of irrigation on yields would not vary with higher order temperature and precipitation effects, nor with the topographical features previously mentioned.

model serves as the basis for our empirical exercise.

3 Data Description

The areal data used in this analysis come from several primary sources. To examine cereal yields, we utilize $5' \times 5'$ grid cell data on global cereal yields in the year 2000 from Monfreda et al. (2008). A nice feature of these yield estimates is that they are formed from national and sub-national sources, providing a higher spatial resolution than previous yield estimates. Initially, these yields were given in tons per hectare, but the figures were re-scaled to pounds per hectare. Data on irrigation, soil carbon density (both at a $5' \times 5'$ grid cell level), and soil pH (at $0.5^\circ \times 0.5^\circ$ resolution) come from FAO-Aquastat, the International Geosphere-Biosphere Program (IGBP), and the International Soil Reference and Information Centre (ISRIC), respectively. The latter two data sets were acquired through the Oak Ridge National Laboratories Distributed Active Archive Center (ORNL-DAAC). Additionally, we utilize $1^\circ \times 1^\circ$ grid cell data on various other spatial, environmental, and economic factors obtained from William Nordhaus's G-Econ dataset (Nordhaus et al., 2006).[5] The yield, irrigation, and soil chemistry data were aggregated to $1^\circ \times 1^\circ$ in order for them to be spatially joined to the G-Econ data applying standard GIS techniques. In total, these data consist of 1,906 observations spanning most of mainland Sub-Saharan Africa.[6]

Note that yields are observed only for a subset of all available grid cells (1,553 observations, or roughly 81 percent). Assuming that all cells with zero yield are the result of farmers explicitly choosing not to plant cereals, this conscious decision could lead to biased OLS estimates. This will be the case, for instance, if farmers in Sub-Saharan Africa systematically choose not to plant cereals in grid cells where they believe yields would be very low. This non-random sampling causes the so-called sample selection problem. The decision to plant cereals is summarized by a binary

[5] The G-Econ data report observations on a country-grid cell level, rather than simply at a grid cell level. For this reason, there are many coordinates that are entered more than one time in the data. We focus on unique grid cell observations, choosing to use the observation that has the largest RIG (i.e., a "rate in grid" observation for each country) measure among all observations for a particular set of geographic coordinates.

[6] In the econometric analysis, Zimbabwe and Somalia were removed from the sample due to missing data on some key explanatory variables. In addition, we omit Madagascar and smaller island states such as Cape Verde, Comoros, Mauritius, Réunion, St. Helena, Seychelles, and São Tomé & Príncipe. The Gambia was omitted as well, since observations for the Gambia consistently had smaller RIG measures than observations for Senegal that shared the same geographic coordinates.

variable taking the value of 1 if cereals were planted in the grid cell in question, and 0 otherwise. A parametric two-step estimation procedure proposed by Heckman (1976, 1979) is widely used to address the sample selection problem. In specifying such a model, however, we need at least one variable that explains this decision to plant cereals while not affecting the resulting yields, which provides an exclusion restriction by which the model can be identified. Our identification strategy therefore uses a grid cell Human Influence Index (HII) to satisfy the exclusion restriction. The HII is an index created by the Socioeconomic Data and Applications Center (SEDAC) of the Center for International Earth Science Information Network (CIESIN) at Columbia University. These data, originally reported at 30 arc second $\times$ 30 arc second grid cell level, aim to "represent the location of various factors presumed to exert an influence on ecosystems: human population distribution, urban areas, roads, navigable rivers, and various agricultural land uses" (CIESIN, 2011). Because this index captures factors such as transportation infrastructure and population density, it is likely to affect the decision to plant cereals without affecting the resulting yields.

Summary statistics for these data can be found in Table 1. The spatial distribution of yields, temperature, precipitation, and the standard deviations of temperature and precipitation as well as the Human Influence Index across Sub-Saharan Africa can be seen in Figure 1, panels (a) through (f), respectively.

[Table 1 about here]

[Figure 1 about here]

Generally, for the purposes of examining spatial effects, grid cell data can be thought of as a regular lattice. From this regular lattice, it becomes a very simple procedure to construct neighborhood structures and weights matrices based on simple contiguity of either the rook or queen type.[7] Modeling spatial systems with sample selection generally results in irregularly distributed observations, since there are cells for which the dependent variable is unobserved. For this reason, we are unable to model a strictly contiguous spatial system. Nevertheless, because the data points

[7]These forms of contiguity take their name from the game of chess, in which rooks can only move in the vertical and horizontal direction, whereas the queen can move vertically and horizontally, as well as along any of the diagonals.

are the geographical centroids of the grid cells, one can *approximate* a neighborhood structure (and thus a row-standardized weights matrix) based on contiguity if one uses a distance-based neighborhood system in which the distance specified is the minimum distance required to ensure that each observation has at least one neighbor.

4 Econometric Estimation and Results

We begin by imposing a specific restriction on this general model, namely the absence of spatial autocorrelation. In the absence of spatial correlation, $\rho = 0$ so the model reduces to:

$$y_i = \mathbf{x}_i'\beta + \mathbf{h}_i'\pi + (\mathrm{Irr}_i \cdot \tilde{\mathbf{x}}_i)'\,\xi + u_i \tag{3}$$

which can be re-written in matrix notation as:

$$y = \mathbf{X}\beta + \mathbf{H}\pi + \left(\mathrm{Irr}' \cdot \tilde{\mathbf{X}}\right)\xi + u$$

Under the assumptions of spherically distributed error terms, this equation can be estimated consistently using ordinary least squares (OLS). A Breush-Pagan test with random coefficients as the alternative hypothesis indicates that the null hypothesis of homoskedasticity should be rejected (the test statistic is 135.65 with a p-value of 9.23×10^{-9}). We therefore report White-adjusted standard errors as well in Table 2.

[Table 2 about here]

From these estimates we can compute the elasticity of yields with respect to temperature and precipitation:

$$\frac{\partial \mathrm{Yield}}{\partial \mathrm{Temperature}} \cdot \frac{1}{\mathrm{Yield}} = \frac{\beta_{\mathrm{Temp}} + 2 \cdot \beta_{\mathrm{Temp}^2} \cdot \mathrm{Temperature}}{\mathrm{Yield}}$$

and

$$\frac{\partial \text{Yield}}{\partial \text{Precipitation}} \cdot \frac{\text{Precipitation}}{\text{Yield}} = \beta_{\text{Precip}} + 2 \cdot \beta_{\text{Precip}^2} \cdot \text{Precipitation} \cdot \frac{\text{Precipitation}}{\text{Yield}}$$

where these elasticities are computed for each observation, and then averaged in order to generate an interpretable figure. Based on these calculations, we see that the semi-elasticity of cereal yields with respect to temperature is –7.74, implying that a 1° C increase in temperature lowers cereal yields in Sub-Saharan Africa by about 7.74 percent on average, and this estimated elasticity is statistically different from zero. The elasticity of yields with respect to precipitation is 10.62, although it should be noted that this elasticity estimate is not statistically different from zero. A larger standard deviation in both temperature and precipitation raises yields at rates that are each statistically different from zero. Since the precipitation measures represent an average monthly rainfall amount, these estimates suggest that a greater dispersion of rainfall is advantageous for cereal yields. This is consistent with literature suggesting that temporal concentration of rainfall is preferred, with greater rainfall desirable in growing seasons and less desirable in harvest seasons.

We proceed by estimating a model that controls for sample selection, following Heckman's parametric two-step approach. In the first stage of this procedure, a linear selection equation of the form $P_i = \mathbf{z}_i'\alpha + \mu_i, \;\; P_i \in \{0, 1\}$ is estimated using a probit estimator, and observation-specific estimates of the inverse Mills ratio (IMR) are constructed as $\lambda_i = \phi(\mathbf{z}_i'\alpha)/\Phi(\mathbf{z}_i'\alpha)$, where $\phi(\cdot)$ is the normal probability density function and $\Phi(\cdot)$ is the normal cumulative distribution function. In the second stage, the outcome (or response) equation for those observations for which $P_i = 1$ is estimated via OLS, with the IMR included as an additional explanatory variable that accounts for the sample selection bias. The conditional response equation is therefore:

$$E[y_i|P_i = 1, \mathbf{x}_i, \tilde{\mathbf{x}}_i, \mathbf{z}_i] = \mathbf{x}_i\beta + \mathbf{h}_i'\pi + (\text{Irr}_i \cdot \tilde{\mathbf{x}}_i)'\,\xi + \eta \left[\frac{\phi(\mathbf{z}_i'\alpha)}{\Phi(\mathbf{z}_i'\alpha)}\right] \tag{4}$$

As previously discussed, specifying such a model requires at least one variable that explains the selection to plant cereals while not affecting the resulting yields, thereby providing an exclusion restriction by which the model can be identified. Our identification strategy therefore uses the grid cell HII variable. While the inclusion of the HII in the selection equation aids in identifying the parameters of the outcome equation, our estimation approach also involves omitting the country dummy variables and the interaction terms from the set of variables conditioning the selection to plant cereals.[8] Estimates of the model parameters computed via Heckman's two-step estimation procedure are presented in column (a) of Table 3.

[Table 3 about here]

From the top panel of column (a) in Table 3, we see that most of the exogenous variables assumed to affect yields also affect the *a priori* decision to plant cereals. The probit model does a very good job of predicting which grid cells will have cereals planted (correctly predicting 98 percent of these "successes"), and similarly does a fair job predicting which grid cells will not have cereals planted (correctly predicting 90 percent of these "failures"). For covariates included in both the selection equation and the outcome equation, the estimates in the lower panel of Table 3 do not represent marginal effects, since the conditional expectation of cereal yields conditional on the explanatory variables should also incorporate the inverse Mills ratio. Consider a covariate included linearly in both the selection and outcome equations, $x_j \in \mathbf{x}, \mathbf{z}$. The marginal effect of this covariate can be computed as:

$$\frac{\partial E[y_i|P_i = 1, \mathbf{x}_i, \tilde{\mathbf{x}}_i, \mathbf{z}_i]}{\partial x_{ij}} = \beta_j - \eta(\alpha_j) \left[(\lambda_i)^2 + \left(z_i' \alpha \right) \lambda_i \right] \tag{5}$$

and subsequently, we can compute semi-elasticities and elasticities. We estimate a semi-elasticity of yields with respect to temperature of –7.58, suggesting that a 1°C increase in temperature would be expected to lower yields by 7.58 percent on average. After controlling for the effect of temperature on the decision to plant cereals in the selection equation, we find that increasing temperatures have a more muted effect on yields than when the selection issue is ignored. The statistically significant coefficient associated with the IMR shows that sample selectivity is a problem, which suggests that

[8] While it may seem plausible that the country dummy variables captures effects that would be relevant in predicting farmer's decisions to plant cereals in a particular grid cell, there are several countries in which all grid cells have cereals planted. Thus the binary country indicator variables for these countries would completely determine the binary selection variable. For this reason, the set of country dummy variables were excluded from the selection equation.

previous cross-sectional studies that do not consider selectivity inconsistently estimate the effects of climatological and other explanatory factors on yields. The estimated elasticity of yields with respect to temperature is lower under the heckit model specification. We find the elasticity of cereal yields with respect to precipitation to be 2.43, which is smaller than the estimated elasticity from the OLS regression.

While this model *does* control for sample selection bias, it does not control for spatial correlation in unobserved factors. To allow for these correlations, we employ an estimator for spatial error models that controls for endogenous sample selection, recently developed by Flores-Lagunes and Schnier (2010). The estimator that Flores-Lagunes and Schnier (2010) propose maintains the intuition of Heckman's model, but is within the broader family of GMM estimators. Specifically, it uses a selection equation analogous to the spatial probit estimator of Pinkse and Slade (1998). The estimates from the spatial probit are then used to construct the spatial econometric equivalent of the IMR, which is then included in the outcome equation. The Pinkse and Slade estimator yields consistent estimates of the selection equation, which are themselves necessary to obtain consistent estimates of the parameters in the response equation. Flores-Lagunes and Schnier (2010) note that when the parameters in the selection equation are different from those in the outcome equation (as is generally desirable to ensure identification), the appropriate IMR is a function of the spatial correlation coefficient in the outcome equation. To increase the efficiency of the estimates, all of the model parameters are estimated simultaneously through solving a system of stacked moment conditions. While this estimator is less efficient than a maximum likelihood estimator, it is consistent and asymptotically normally distributed with an estimable variance-covariance matrix.

To proceed, let yields be represented as a latent variable y_i^* and the planting decision be represented by the latent variable P_i^*, such that:

$$\begin{aligned} &y_i = y_i^* \quad \text{if } P_i^* > 0 \quad y_i = 0 \text{ otherwise} \\ &P_i = 1 \quad \text{if } P_i^* > 0 \quad P_i = 0 \text{ otherwise} \end{aligned}$$

Explicitly modeling the selection and response equations taking into consideration spatial dependence in the errors, we have $P_i^* = \mathbf{z}_i'\alpha + \varepsilon_{1i}$, $\varepsilon_{1i} = \rho_1 \sum_{j \neq i} \mathrm{w}_{ij}^1 \varepsilon_{1j} + v_i$ and $y_i^* = \mathbf{x}_i'\beta + \mathbf{h}_i'\pi +$

$(\mathrm{Irr}_i \cdot \tilde{\mathbf{x}}_i)' \xi + \varepsilon_{2i}$, $\varepsilon_{2i} = \rho_2 \sum_{j \neq i} \mathrm{w}^2_{ij} \varepsilon_{1j} + u_i$, where w^1_{ij} is the (i,j) element of the $(n+b) \times (n+b)$ spatial weights matrix corresponding to the selection equation, and w^2_{ij} is the (i,j) element of the $n \times n$ spatial weights matrix corresponding to the outcome equation. Note that both equations are general enough to allow for spatial dependence in the errors, where the degree of spatial correlation is denoted by coefficients ρ_1 and ρ_2, respectively. The innovations v and u are assumed *iid* and multivariate normal such that $(v_i, u_i) \sim N(0, \Sigma)$, where:

$$\Sigma = \begin{bmatrix} \sigma_v^2 & \sigma_{vu} \\ \sigma_{vu} & \sigma_u^2 \end{bmatrix}$$

From these equations, we can write the model equations in their reduced form:

$$P_i^* = \mathbf{z}_i' \alpha + \sum_j \omega_{1,ij} v_j \tag{6}$$

$$y_i^* = \mathbf{x}_i' \beta + \mathbf{h}_i' \pi + (\mathrm{Irr}_i \cdot \tilde{\mathbf{x}}_i)' \xi + \sum_j \omega_{2,ij} u_j \tag{7}$$

where $\omega_{k,ij} = \left[(\mathbf{I} - \rho_k \mathbf{W}_k)^{-1}\right]_{ij}$ is the (i,j) element of the $k-$th equation spatial multiplier matrix $(\mathbf{I} - \rho_k \mathbf{W}_k)^{-1}$. Flores-Lagunes and Schnier (2010) note that the probit model with spatially autoregressive errors introduces a fully non-spherical variance-covariance matrix that renders the regular probit estimator inconsistent. Estimation of the selection equation therefore proceeds along the lines of the spatial probit model of Pinkse and Slade (1998), to obtain consistent estimates.

From McMillen (1995), we have the following variance and covariance calculations:

$$\mathrm{Var}(\varepsilon_1) = \sigma_v^2 \sum_j (\omega_{1,ij})^2$$

$$\mathrm{Var}(\varepsilon_2) = \sigma_u^2 \sum_j (\omega_{2,ij})^2$$

$$E(\varepsilon_{1i}\varepsilon_{2i}) = \sigma_{vu} \sum_j \omega_{1,ij}\omega_{2,ij}$$

Using $\mathrm{Var}(\varepsilon_{1i})$, Pinkse and Slade (1998) construct "generalized" residuals with which to construct appropriate moment conditions for consistent estimation of the model parameters, taking into consideration the induced heteroskedasticity. Letting the vector of parameters in the selection equation be given as $\theta_1 = [\alpha', \rho_1]$ and letting $\delta_i(\theta_1) = \mathbf{z}_i'\alpha/\sqrt{\mathrm{Var}(\varepsilon_{1i})}$ be the index function of a probit model weighted by the standard deviation of the residuals from the selection equation, the "generalized" residuals of the selection equation are:

$$\tilde{\varepsilon}_{1i}(\theta_1) = \sqrt{\sigma_v^2 \sum_j (\omega_{1,ij})^2} \cdot \{P_i - \Phi[\delta_i(\theta_1)]\} \cdot \frac{\phi[\delta_i(\theta_1)]}{\Phi[\delta_i(\theta_1)]\{1-\Phi[\delta_i(\theta_1)]\}}$$

The GMM estimator for θ_1 is given as:

$$\theta_{1,GMM} = \mathrm{argmin}\left\{S(\theta_1)'\mathbf{M}_n S(\theta_1)\right\} \tag{8}$$

where $S(\theta_1) = \frac{1}{n}\mathbf{Z}'\tilde{\varepsilon}_1(\theta_1)$, where $\mathbf{Z}$ is the matrix of variables in the selection equation and $\tilde{\varepsilon_1}(\theta_1)$ is the vector of generalized residuals, and $\mathbf{M}_n$ is a conformable positive definite moment-weighting matrix. Consistent estimates of θ_1 are then used to construct the "adjusted" IMR (McMillen, 1995) to be used in the outcome equation. The "adjusted" IMR is given as:

$$\hat{\lambda}_i \equiv \frac{\sum_j \omega_{1,ij}\omega_{2,ij}}{\sqrt{\sum_j (\omega_{1,ij})^2}} \cdot \frac{\phi[-\delta_i(\theta_1)]}{1-\Phi[-\delta_i(\theta_1)]} \tag{9}$$

This "adjusted" IMR depends on the spatial correlation coefficient from the outcome equation (ρ_2), which is not estimated in the first-stage spatial probit. Likewise estimating the conditional outcome response (including estimating the spatial correlation coefficient) requires the inclusion of this "adjusted" IMR as an additional explanatory variable in the outcome equation. All of the parameters in both the selection and the outcome equations can be estimated simultaneously in order to increase the efficiency of the estimator. To accomplish this, Flores-Lagunes and Schnier

(2010) stack the moment conditions of the selection and outcome equations:

$$g(\mathbf{Z}, \mathbf{X}, \mathbf{H}, \tilde{\mathbf{X}}, \theta) = \left[s(\mathbf{Z}, \theta)', m(\mathbf{X}, \mathbf{H}, \tilde{\mathbf{X}}, \theta)'\right]'$$

where, now, $\theta = [\alpha', \rho_1, \beta', \pi', \xi', \eta, \rho_2]$. The components of this matrix are as follows:

$$s(\mathbf{Z}, \theta) = \mathbf{Z}'\tilde{\varepsilon_1}(\theta)$$

$$m(\mathbf{X}, \mathbf{H}, \tilde{\mathbf{X}}) = \left[P \cdot \left(\mathbf{X}, \mathbf{H}, \tilde{\mathbf{X}}, \hat{\lambda}\right)\right]' \tilde{\varepsilon}_2(\theta)$$

where $\tilde{\varepsilon}_1$ is the vector of generalized residuals and $\tilde{\varepsilon}_2(\theta) = y - \mathbf{X}\beta - \mathbf{H}\pi - (\text{Irr} \cdot \tilde{\mathbf{X}}) - \eta\hat{\lambda}(\rho_1, \rho_2, \alpha)$. Stacking the generalized residuals from the spatial probit estimation with the residuals from the outcome equation, we get $\tilde{\varepsilon}(\theta) \equiv [\tilde{\varepsilon}_1'(\theta), \tilde{\varepsilon}_2'(\theta)]$. Then a consistent GMM estimator for all of the model parameters is:

$$\theta_{GMM} = \text{argmin}\left\{g_n(\theta)'\mathbf{M}_n g_n(\theta)\right\} \tag{10}$$

where $g_n = \dfrac{1}{n}\mathbf{Z}'\tilde{\varepsilon}(\theta)$ and, again, $\mathbf{M}_n$ is a conformable positive definite moment-weighting matrix. We use the optimally-weighted GMM estimator, which obtains efficient estimates of the model parameters in two steps. The first step estimates the model parameters using an equally-weighted GMM estimator, from which the optimal weights are constructed. These optimal weights are then used in the second step GMM estimation.

The results from the spatial heckit estimation of equations (6) and (7) are reported in column (b) of Table 3. There is highly significant (and strong) spatial correlation among the unobserved error terms in both the selection and response equations. The spatial correlation coefficient in the selection equation is 0.62, while the spatial correlation in the outcome equation is 0.69, suggesting the potential for a high degree of spill-overs among neighboring grid cells. Since country fixed effects are not controlled for in the selection equation, it is possible that these unobserved factors capture effects that would otherwise be captured by country effects. In the outcome equation, however, country effects are explicitly controlled for, so the unobserved errors in the outcome equation must be distinct from country effects.

The estimates in the lower panel of Table 3 cannot be interpreted as *marginal effects* of small changes in the explanatory variable on cereal yields.[9] Interpreting these coefficient estimates as such ignores some potentially important nonlinear effects introduced through the incorporation of the "adjusted" IMR as an additional explanatory variable in the outcome equation. Thus the true marginal effects should take into consideration both the *direct* effect that the explanatory variable in question has on the outcome variable, and also, for particular explanatory variables, the *indirect* effect of these independent variables on the probability that the selection equation dependent variable is positive. Expected yields conditional upon the affirmative decision to plant cereals are given as:

$$E[y_i|P_i^* > 0, \mathbf{x}_i, h_i, \tilde{\mathbf{x}}_i] = \mathbf{x}_i'\beta + \mathbf{h}_i'\pi + (\mathrm{Irr}_i \cdot \tilde{\mathbf{x}}_i)'\,\xi + \eta\hat{\lambda}_i \tag{11}$$

This expression can be expanded and re-written as:

$$E[y_i|P_i^* > 0, \mathbf{x}_i, h_i, \tilde{\mathbf{x}}_i] = \mathbf{x}_i'\beta + \mathbf{h}_i'\pi + (\mathrm{Irr}_i \cdot \tilde{\mathbf{x}}_i)'\,\xi + \eta \cdot \left(\frac{\displaystyle\sum_j \omega_{ij}^1 \omega_{ij}^2}{\sqrt{\displaystyle\sum_j \left(\omega_{ij}^1\right)^2}} \cdot \frac{\phi\left[-\frac{\mathbf{z}_i'\alpha}{\sqrt{\mathrm{Var}(\varepsilon_{1i})}}\right]}{1 - \Phi\left[-\frac{\mathbf{z}_i'\alpha}{\sqrt{\mathrm{Var}(\varepsilon_{1i})}}\right]} \right)$$

Writing the conditional expectation in this form provides some valuable insight into the benefits of modeling spatial dependence. In most applications, the modeling spatial correlation in the unobserved error term simply increases the efficiency of the parameter estimates by correcting bias in the variance-covariance matrix. This expanded expression indicates that the failure to account for spatial correlation when the data exhibit sample selection bias can itself lead to biased parameter estimates, in addition to problems of inefficiency. This is seen by the inclusion of the $\sum_j \omega_{ij}^1 \omega_{ij}^2 / \sqrt{\sum_j \left(\omega_{ij}^1\right)^2}$ term, which incorporates the spatial multiplier matrices $(\mathbf{I} - \rho_k \mathbf{W}_k)^{-1}$,

[9] Clearly, if $\mathbf{x}_j \notin \mathbf{z}$, then small changes in $\mathbf{x}_j$ do not affect the probability of planting cereals, and therefore the marginal effects for such explanatory variables are simply the parameter estimates obtained from the outcome equation.

$k = 1, 2$. If spatial correlation is nonexistent in neither the selection or the outcome equation, then $\rho_1 = \rho_2 = 0$, and $(\mathbf{I} - \rho_k \mathbf{W}_k)^{-1} = \mathbf{I}$, so that $\sum_j \omega^1_{ij}\omega^2_{ij} = \sqrt{\sum_j \left(\omega^1_{ij}\right)^2} = 1 \quad \forall i$, and the conditional expectation obtained from equation (11) is the same as what would be obtained by estimating a sample selection model without incorporating spatial correlation in the error terms. If, however, spatial correlation in the error terms is present, then estimating the model using a standard, non-spatial approach will result in conditional expectations that may be significantly different than those obtained from models incorporating spatial effects. This also has implications for the computation of marginal effects. For a generic linear covariate included in both the selection and outcome equations (i.e., $x_j \in \mathbf{x}$, $\mathbf{z}$), the marginal effects are given by:

$$\frac{\partial E[y_i|P^*_i > 0, \mathbf{x}_i, \mathbf{h}_i, \tilde{\mathbf{x}}_i]}{\partial x_j} = \beta_j - \eta \left(\frac{\alpha_j}{\sqrt{\mathrm{Var}\,(\varepsilon_{1i})}}\right)\left(\frac{\sum_j \omega^1_{ij}\omega^2_{ij}}{\sqrt{\sum_j \left(\omega^1_{ij}\right)^2}}\right)\left[(\lambda_i)^2 + (\delta_i(\theta_1))\,\lambda_i\right] \tag{12}$$

where $\lambda_i \equiv \phi[\delta_i(\theta_1)]/\Phi[\delta_i(\theta)] \neq \hat{\lambda}_i$ is the standard (i.e., not "adjusted") inverse Mills ratio. Again, the marginal effects will be biased if the spatial correlation is not incorporated. Since many of the explanatory variables for which the marginal effects would be particularly interesting enter the outcome equation in a nonlinear fashion, the above general form for computing marginal effects needs to be modified to incorporate higher order effects and interactions.

Because of the inherent nonlinearities introduced through the IMR, the marginal effects will vary by observation. There are generally two standard approaches for estimating average effects. One standard approach involves evaluating the marginal effects at the mean of the data. Another approach involves computing the marginal effect for each observation and then averaging. Greene (2003) has suggested that, assuming the data are well-behaved, these two approaches are asymptotically equivalent. In our analysis, we used the latter approach because there is a great deal of variation in values for the exogenous explanatory variables across Sub-Saharan Africa.

Obviously, the marginal effects, elasticities and semi-elasticities are themselves nonlinear functions of the model parameters. To compute standard errors and confidence intervals for both elasticities and semi-elasticities, we employ an approach analogous to parametric bootstrap techniques introduced by Krinsky and Robb (1986). In this approach, random drawings are taken from a multivariate normal distribution with a mean and variance-covariance matrix obtained in the estimation of the model parameters. For each drawing from this multivariate normal distribution, the elasticities are re-calculated, generating an empirical frequency distribution from which confidence intervals can be obtained. The semi-elasticities and elasticities of yields with respect to temperature and precipitation changes (respectively) are reported in panels (a) and (b) in Table 4 for each of the three econometric models estimated.

[Table 4 about here]

All three models predict roughly the same elasticity of yields with respect to temperature changes, and they are significantly different from zero based on bootstrap confidence intervals. All three models suggest that a 1°C increase in temperatures lowers cereal yields by roughly 7.5 percent. The semi-elasticity computed from the OLS estimates suggest the largest reduction in yields, while the semi-elasticity computed from the non-spatial heckit suggest the smallest reduction in yields among the three models considered. The semi-elasticity computed from the spatial heckit estimator lies within the range spanned by these two non-spatial estimators. The results from the three models predict a positive effect on yields with increased precipitation, although these elasticities are not significantly different from zero. The reason may be because simple changes in monthly mean precipitation are difficult to interpret as causal variables, since the timing of such precipitation is vitally important. As was the case with the temperature semi-elasticities, the precipitation elasticities vary widely over space.

While not reported, it should be noted that many of the coefficients associated with the country dummy variables are significantly different from zero. This suggests that, even after controlling for the effects of temperature, precipitation, and soil chemistry, there is strong evidence that country-specific factors do significantly contribute to predicting cereal yields.

5 Simulation Experiment

Of all the explanatory variables we include in our model, only grid cell irrigation is explicitly under the control of policymakers. Analysis of this variable allows us to investigate how policy interventions (such as improvements to or expansions of irrigation infrastructure) affect cereal yields in Sub-Saharan Africa, both in the absence of any climatic changes and under assumptions of spatially-heterogeneous temperature and precipitation changes drawn from a combination of general circulation models. Because temperature and precipitation are interacted with grid-cell level irrigation in the outcome equation of the spatial heckit model, the effect of changes in these climatic variables on cereal yields will vary over space based on the level of irrigation in a particular grid cell. As would be expected, average yields for grid cells with irrigation (1,129 kg per hectare) are significantly higher than for grid cells without irrigation (773 kg per hectare). Irrigation helps to offset some of the declines in yields brought on by higher temperatures (see the lower panel of column (b) in Table 3). From the spatial heckit model, we see that a 1 percent increase in grid cell access to irrigation increases yields by roughly 1.35 percent, although the estimated elasticity is not significantly different from zero (see panel (c) of Table 4).

[Table 5 about here]

We can use equation (11) to estimate the effects of additional irrigation on expected yields both with and without climate change. Table 5 reports average temperatures and levels of precipitation both historically and in the future for a select group of countries and regions in Sub-Saharan Africa. These figures represent the averages from 23 general circulation models compiled by Cline (2007). Using the projected changes in temperatures and precipitation, we modify our initial data to simulate future climate conditions.[10] Based on these climate conditions, we estimate that average cereal yields across Sub-Saharan Africa will be only 806 kg per hectare, down approximately 15 percent from the currently observed average level of 950 kg per hectare. From this new base

[10]For simplicity, we are assuming that the shape of a grid cell's annual distribution of temperatures and monthly precipitation levels are unchanged, but rather that the distributions are simply shifting laterally. This simplification allows us to assume that the standard deviations of temperature and precipitation will be the same in the future as they are at present.

average yield, we can examine how improvements in irrigation infrastructure can be expected to affect yields, partially ofsetting some of the damage resulting from anticipated climatic changes. We consider four scenarios for infrastructure investment:

(i) Uniform irrigation increase of 5 percentage points

(ii) Targeted increase of 5 percentage points for cells with some (but incomplete) irrigation coverage

(iii) Targeted increase of 5 percentage points for cells with high levels of irrigation

(iv) Targeted increase of 5 percentage points for cells with no irrigation

For scenario (ii), we identify cells where the irrigation coverage is in the bottom quartile of all nonzero levels of irrigation. For this sample, it is the case that roughly 25 percent of grid cells with nonzero levels of irrigation coverage only have irrigation infrastructure in place for 0.65 percent or less of the total area of the grid cell. For scenario (iii), we identified cells with levels of irrigation in the top quartile of cells with positive levels of irrigation, which corresponds with irrigation coverage of only 3.81 percent of the grid cell or more. The results from these various scenarios are summarized in Table 6. The results suggest that increasing irrigation infrastructure uniformly results in the largest average increase in cereal yields, with yields increasing 14 percent. While this may significantly offset many of the expected losses resulting from climate change, this strategy is also not very feasible, since it would require a massive mobilization of resources. Excluding this scenario, the largest increase in yields is derived by implementing irrigation where there currently is none. Under this scenario, average yields are expected to be 7.6 percent higher than the base scenario incorporating climate change. While the yield increases resulting from this strategy are not nearly as significant as when the irrigation increases are across-the-board, the results of this simulation suggest that targeting those cells without any existing irrigation infrastructure should provide a larger average response than other targeting strategies.

[Table 6 about here]

6 Conclusion

In this paper we estimated a spatial econometric model to examine grid cell level cereal grain productivity in Sub-Saharan Africa. Our model specification allowed us to examine the dependence of agricultural productivity on a variety of exogenous climatological and topographical factors, taking into consideration spatial dependence and spatial heterogeneity while controlling for sample selection. Consistent with many previous studies, we find that increasing temperatures have a negative effect on cereal yields. Regardless of whether spatial dependence is modeled, our results suggest that failing to control for sample selection bias may lead to overestimation of the negative impacts of increasing temperatures on cereal yields. We cannot draw any conclusions about how this omission affects estimates of the effects of precipitation on yields, since the resulting elasticities with respect to precipitation are not significantly different from zero. Our findings suggest that there is considerable spatial dependence affecting both the cell-level decision to plant cereals as well as the conditional yield response.

While not reported, we find strong statistical evidence to support the notion that, even after controlling for temperature, precipitation and soil characteristics, effects at the country level are correlated with yields. Binary variables corresponding to a cell's country of origin capture country-specific effects, which are highly statistically significant and act as yield shifters, suggesting substantial spatial heterogeneity across Sub-Saharan Africa. These country effects could incorporate many factors, including governmental subsidies for fertilizers, the general availability of fertilizers, extension programs to assist farmers' decision-making, a sound legal system which provides stringent property rights, a system of functioning input and output markets, supply chain management, price transmission, etc. We do not attempt to disentangle the sources of these country effects in this study, leaving that task to future research.

Our methodological approach attempts to blend the spatially-explicit nature of agronomic crop growth models with the simplicity of statistical models. By using spatially-explicit data at a relatively high spatial resolution, we are able to partially avoid ecological biases that are often present when estimating such relationships at coarser resolutions. We also incorporate a recently developed estimator for spatial process models that controls for sample selection bias. While the use of $1^\circ \times 1^\circ$

grid cell data allows for interpretation of marginal effects at a higher resolution than the use of country-level data would allow, our estimates are still partially subject to ecological bias, since farms are ultimately the decision-making units of interest. Additionally, this approach does not explicitly allow for adaptation, which is one of the touted advantages of the Ricardian approach. Similarly, unlike the agronomic crop yield studies or other process based approaches, our approach quantifies only statistical correlations, without necessarily identifying causality. While this may technically be a weakness of our methodological approach, the strict exogeneity of many of our explanatory variables presumes a very direct causal interpretation. Data limitations preclude modeling yield responses in a manner more amenable to agronomic studies. We only have observations on annual average temperatures and average monthly precipitation, whereas ideally one would prefer to have average growing season temperatures or total temperature days and average growing season precipitation. We also lack observations on other traditional inputs into agricultural production, such as labor hours and fertilizer applications. Additionally, because this study focuses on a cross-sectional analysis, we forgo any dynamical elements that affect crop yields. If we assume trend stationarity, then it could plausibly be argued that cross-sectional differences in yields could capture some of the dynamical or temporal differences as well. The issue of how representative cross-sectional differences are of dynamic differences over the ensuing decades warrants future investigation.

The elasticities estimated using simple OLS and the much more complex spatial sample selection estimator are not vastly different from one another, especially for the semi-elasticities of yields with respect to temperature and irrigation. This may lead some observers to question the value of such a complex estimator. We argue that, while the benefits of this complexity may not be readily apparent from the current study, it represents only one application, and judging the value of a methodological approach based on a sample of one is not justified. We have shown that parameter estimates, marginal effects and elasticities obtained by OLS will be biased under very general violations of the standard OLS assumptions, and that this bias is increasing in the absolute value of both the spatial effect and the sample selection effect. In our particular study, the spatial effect is quite large, but this effect is muted by a relatively small sample selection effect. In applications where these effects are larger, the potential bias correction obtained by using this new methodological approach can be significant. In any event, irrespective of the seemingly small differences between the various elasticity estimates, we highlight the increased accuracy and precision that these new estimates provide, since incorporating spatial dependence and sample selection corrections reduce bias in the estimates themselves as well as in the variance of these estimates.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Springer.

Anselin, L., A. Bera, R. Florax, and M. Yoon (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics 26*(1), 77–104.

CIESIN (2011). Global Human Influence Index. Data set available at: http://sedac.ciesin.columbia.edu/wildareas/downloads.jsp. Accessed 8 February 2011.

Cline, W. (1996). The impact of global warming of agriculture: comment. *The American Economic Review 86*(5), 1309–1311.

Cline, W. (2007). *Global Warming and Agriculture.* Center for Global Development and Peterson Institute for International Economics.

Darwin, R., M. Tsigas, J. Lewandrowski, and A. Raneses (1995). World agriculture and climate change: Economic adaptations. *Agricultural Economics Reports*.

Deaton, A. (1997). *The analysis of household surveys: a microeconometric approach to development policy.* World Bank Publications.

FAO (2011). FAOSTAT crop production data. Available online at http://faostat.fao.org/. Data accessed 23 August 2011.

Flores-Lagunes, A. and K. Schnier (2010). Estimation of sample selection models with spatial dependence. *Journal of Applied Econometrics*, (Forthcoming). doi:10.1002/jae.1189.

Greene, W. (2003). *Econometric analysis* (5 ed.). Upper Saddle River, NJ: Prentice Hall.

Haining, R. (2003). *Spatial data analysis: theory and practice.* Cambridge Univ Pr.

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and social Measurement 5*(4), 475–492.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Hertel, T. and S. Rosch (2010). Climate Change, Agriculture, and Poverty. *Applied Economic Perspectives and Policy 32*(3), 355.

Jones, C., P. Dyke, J. Williams, J. Kiniry, V. Benson, and R. Griggs (1991). Epic: an operational model for evaluation of agricultural sustainability. *Agricultural Systems 37*(4), 341–350.

Jones, C. and J. Kiniry (1986). Ceres-maize: A simulation model of maize growth and development.

Jones, J., G. Hoogenboom, C. Porter, K. Boote, W. Batchelor, L. Hunt, P. Wilkens, U. Singh, A. Gijsman, and J. Ritchie (2003). The dssat cropping system model* 1. *European journal of agronomy 18*(3-4), 235–265.

Kauffmann, D., A. Kraay, and M. Mastruzzi (2010). The worldwide governance indicators : A summary of methodology, data and analytical issues. World Bank Policy Research.

Kelejian, H. and I. Prucha (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics 17*(1), 99–121.

Kelejian, H. and I. Prucha (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics 157*(1), 53–67.

Krinsky, I. and A. Robb (1986). On approximating the statistical properties of elasticities. *The Review of Economics and Statistics 68*(4), 715–719.

Lobell, D. and M. Burke (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*.

Lobell, D., M. Burke, C. Tebaldi, M. Mastrandrea, W. Falcon, and R. Naylor (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science 319*(5863), 607.

McMillen, D. (1995). Selection Bias in Spatial Econometric Models. *Journal of Regional Science 35*(3), 417–436.

Mendelsohn, R. and A. Dinar (1999). Climate change, agriculture, and developing countries: does adaptation matter? *The World Bank Research Observer 14*(2), 277.

Mendelsohn, R., W. Nordhaus, and D. Shaw (1994). The impact of global warming on agriculture: a Ricardian analysis. *The American Economic Review 84*(4), 753–771.

Monfreda, C., N. Ramankutty, and J. Foley (2008). Farming the Planet 2: Geographic Distribution of Crop Areas, Yields, Physiological Types, and Net Primary Production in the Year 2000. *Global Biogeochemical Cycles 22*, GB1022, doi:10.1029/2007GB002947.

Nordhaus, W., Q. Azam, D. Corderi, K. Hood, N. M. V., M. Mohammed, A. Miltner, and J. Weiss (2006). The G-Econ Database on Gridded Output: Methods and Data. Technical report, G-Econ Project, Yale University, New Haven, CT, USA.

Pinkse, J. and M. Slade (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics 85*(1), 125–154.

Rowhani, P. and N. Ramankutty (2009). Crop modeling. Paper presented at the Workshop on Climate Change, Agricultural Variability and Poverty Vulnerability, October 2009. Washington, D.C.: The World Bank. http://siteresources.worldbank.org/INTIE/Resources/mornin2_Rowhani.pdf. Accessed 15 December 2010.

Schlenker, W., W. Hanemann, and A. Firsher (2006). The impact of global warming on U.S. agriculture: an econometric analysis of optimal growing conditions. *Review of Economics and Statistics 88*(1), 113–125.

Schlenker, W., W. Hanemann, and A. Fisher (2005). Will US agriculture really benefit from global warming? Accounting for irrigation in the hedonic approach. *American Economic Review 95*(1), 395–406.

Schlenker, W. and M. Roberts (2006). Nonlinear effects of weather on corn yields. *Applied Economic Perspectives and Policy 28*(3), 391.

WDR (2010). *World Development Report 2010: Development and Climate Change*. The World Bank.

World Bank (2006). *World Development Indicators*. The World Bank Washington, DC.

Table 1: Summary Statistics for Dependent and Independent Variables

Variable	Description	N	Mean	Std. Dev.	Min.	Max.
Yield	Cereal yields, 2000 (pounds per hectare, 000s)	1,553	2.10	1.08	0.16	10.36
Plant	Binary variable (=1 if cereals are planted)	1,906	0.81	0.39	0	1
HII	Human Influence Index	1,906	13.04	7.12	0.00	41.00
Temperature	Average annual temperature (degrees Celsius)	1,906	24.60	3.47	10.21	30.32
Std. Dev. Temp.	Standard deviation of temperature	1,906	2.83	1.86	0.27	7.72
Precipitation	Average monthly precipitation (mm)	1,906	65.58	48.29	0.08	222.69
Std. Dev. Precip.	Standard deviation of precipitation	1,906	55.20	35.76	0.13	246.57
Elevation	Elevation (m)	1,906	677.22	432.97	4.59	2,575.22
Rough.	Roughness of terrain	1,906	8.75	9.96	0.00	60.00
Dist.	Distance to coast (km)	1,906	721.58	436.83	4.10	1,686.70
Irrigation	Irrigation (% of grid cell with improved irrigation)	1,906	1.37	3.73	0.00	52.89
Soil pH	Soil pH index (ranging from 0–99)	1,906	33.48	12.91	15.06	99.00
Soil Carbon	Soil carbon density (kg C/m^2)	1,906	8.02	3.82	0.00	27.72

Table 2: Selected Coefficients: OLS Estimation

	Estimate	**Std. Error**	**Robust Std. Error**
Constant	3.5148	1.1488***	1.4241**
Temperature	−0.0206	0.0881	0.1208
Temp.^{2}	−0.0021	0.0019	0.0029
Std. Dev. Temp.	0.1214	0.0409***	0.0826
Precipitation	0.0017	0.0032	0.0032
Precip.^{2} ($\times 1,000$)	0.0017	0.0128	0.0122
Std. Dev. Precip.	0.0037	0.0014***	0.0013***
Elevation	−0.0004	0.0001***	0.0002*
Roughness	−0.0003	0.0001***	0.0029
Distance to Shore	0.0047	0.0027*	0.0001***
Irrigation	−0.0901	0.0861	0.1184
Soil pH	0.0002	0.0032	0.0028
Soil Carbon	0.0196	0.0064***	0.0050***
Irr. · Temp.	0.0037	0.0020*	0.0025
Irr. · Std. Dev. Temp.	0.0266	0.0079***	0.0100***
Irr. · Precip.	0.0003	0.0005	0.0006
Irr. · Std. Dev. Precip.	0.0002	0.0004	0.0004
Irr. · pH	−0.0007	0.0009	0.0010
Irr. · Carbon	−0.0067	0.0015***	0.0020***
Adjusted $R^2 = 0.61$			
$N = 1{,}553$			

Note: * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level

Table 3: Selected Coefficients: Non-Spatial Heckit and Spatial Heckit Estimation

Selection Equation	(a) Non-Spatial Heckit		(b) Spatial Heckit	
	Estimate	Std. Error	Estimate	Std. Error
Constant	17.9100	3.8650***	18.1188	6.9954***
Human Influence Index	0.0528	0.0162***	0.0558	0.0242**
Temperature	−1.7670	0.3281***	−1.7619	0.5374***
Temp.^{2}	0.0372	0.0068***	0.0374	0.0106***
Std. Dev. Temp.	0.1669	0.0876*	0.1750	0.1212
Precipitation	0.0483	0.0125***	0.0485	0.0136***
Precip.^{2} ($\times 1,000$)	−0.3028	0.0608***	−0.3162	0.0629***
Std. Dev. Precip.	0.0479	0.0071***	0.0435	0.0072***
Elevation	0.0009	0.0005**	0.0009	0.0005*
Distance to Shore	−0.0132	0.0111	−0.0006	0.0003**
Roughness	−0.0006	0.0002***	−0.0136	0.0117
Irrigation	0.1256	0.0300***	0.1265	0.1347
Soil pH	−0.0149	0.0064**	−0.0149	0.0058**
Soil Carbon	0.0899	0.0333***	0.0887	0.0338***
ρ			0.6248	0.0463***
N	1,906		1,906	
Percent "Successes" Correctly Predicted:	97.94%		98.91%	
Percent "Failures" Correctly Predicted:	89.52%		81.87%	
Total Percent Correctly Classified:	96.38%		95.75%	

Outcome Equation	(a) Non-Spatial Heckit		(b) Spatial Heckit	
	Estimate	Std. Error	Estimate	Std. Error
Constant	3.7760	1.1250***	3.7607	1.2837***
Temperature	−0.0792	0.0872	−0.0791	0.0979
Temp^2	−0.0006	0.0019	−0.0006	0.0022
Std. Dev. Temp.	0.0330	0.0408	0.0337	0.0612
Precipitation	0.0044	0.0032	0.0045	0.0032
Precip.^{2} ($\times 1,000$)	−0.0273	0.0134**	−0.0270	0.0159*
Std. Dev. Precip.	0.0080	0.0015***	0.0081	0.0016***
Elevation	−0.0003	0.0001**	−0.0003	0.0002*
Roughness	0.0038	0.0027	−0.0004	0.0001***
Distance to Shore	−0.0004	0.0001***	0.0038	0.0026
Irrigation	−0.0905	0.0841	−0.0904	0.1066
Soil pH	−0.0016	0.0031	−0.0016	0.0027
Soil Carbon	0.0269	0.0066***	0.0269	0.0047***
Irr. · Temp.	0.0050	0.0020**	0.0050	0.0022**
Irr. · Std. Dev. Temp.	0.0273	0.0075***	0.0280	0.0104***
Irr. · Precip.	0.0005	0.0005	0.0006	0.0005
Irr. · Std. Dev. Precip.	−0.0002	0.0005	−0.0002	0.0004
Irr. · pH	−0.0009	0.0009	−0.0011	0.0009
Irr. · Carbon	−0.0079	0.0015***	−0.0079	0.0018***
Inverse Mills Ratio	0.8844	0.0863***		
Adjusted Inverse Mills Ratio			0.8788	0.3522**
ρ			0.6926	0.0281***
N	1,553		1,553	
Adjusted R^2	0.63			

Note: * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level

Table 4: Elasticities of Yields With Respect to (a) Temperature, (b) Precipitation, and (c) Irrigation

	Estimate	95% Confidence Interval Lower 2.5% Tail	95% Confidence Interval Upper 2.5% Tail	Mean from 1,000 Replications
(a)				
OLS	–7.74	–11.08	–4.65	–7.71
Heckit	–7.58	–11.01	–4.86	–7.70
Spatial Heckit	–7.62	–11.87	–3.40	–7.62
(b)				
OLS	10.62	-1.89	23.41	10.77
Heckit	2.43	-10.55	13.82	2.10
Spatial Heckit	5.28	-8.30	18.51	5.21
(c)				
OLS	1.43	0.36	2.51	1.44
Heckit	1.17	–0.14	2.40	1.14
Spatial Heckit	1.35	–0.65	3.28	1.31

Table 5: Present and Future Climate: Average Temperature and Precipitation

	Temperature (°C)		Change	Precipitation (mm per day)		Change
	1961-90	2070-99	(°C)	1961-90	2070-99	(%)
Angola	21.52	25.53	4.01	2.75	2.62	-4.73%
Burkina Faso	28.16	32.38	4.22	2.12	2.29	8.02%
Cameroon	24.60	28.16	3.56	4.36	4.50	3.21%
DR Congo	23.95	27.93	3.98	4.21	4.27	1.43%
Ethiopia	23.08	26.92	3.84	2.04	1.97	-3.43%
Ghana	27.15	30.87	3.72	3.23	3.27	1.24%
Cote D'Ivoire	26.19	29.79	3.60	3.88	3.95	1.80%
Kenya	24.33	27.83	3.50	2.02	2.19	8.42%
Madagascar	22.28	25.53	3.25	4.12	3.91	-5.10%
Malawi	21.79	25.72	3.93	3.10	3.04	-1.94%
Mali	28.24	33.01	4.77	0.85	0.87	2.35%
Mozambique	23.44	27.28	3.84	2.82	2.80	-0.71%
Niger	27.13	31.53	4.40	0.46	0.68	47.83%
Nigeria	26.73	30.46	3.73	3.09	3.29	6.47%
Senegal	27.80	31.51	3.71	1.95	1.80	-7.69%
South Africa	17.72	21.89	4.17	1.31	1.20	-8.40%
Sudan	26.70	30.87	4.17	1.18	1.28	8.47%
Tanzania	22.25	26.01	3.76	2.88	2.91	1.04%
Uganda	22.36	26.04	3.68	3.24	3.30	1.85%
Zambia	21.57	25.86	4.29	2.75	2.61	-5.09%
Zimbabwe	21.03	25.39	4.36	1.85	1.81	-2.16%
Other Equatorial Africa	24.81	28.46	3.65	4.23	4.30	1.65%
Other Horn of Africa	26.79	30.35	3.56	0.81	0.96	18.52%
Other West Africa	25.77	29.29	3.52	5.24	5.32	1.53%
Average			3.93			2.52%

Source: Cline (2007)

Table 6: Results of Simulation Experiments: Yield Changes Resulting from Improvements to Irrigation Infrastructure

Scenario	Change in Yields
(i)	14.3%
(ii)	3.4%
(ii)	3.7%
(iv)	7.6%

Note: Yield changes reflect average changes in cereal yields resulting from improvements to irrigation infrastructure compared to a base scenario incorporating spatially-heterogeneous temperature and precipitation changes based off estimates reported in Table 5.

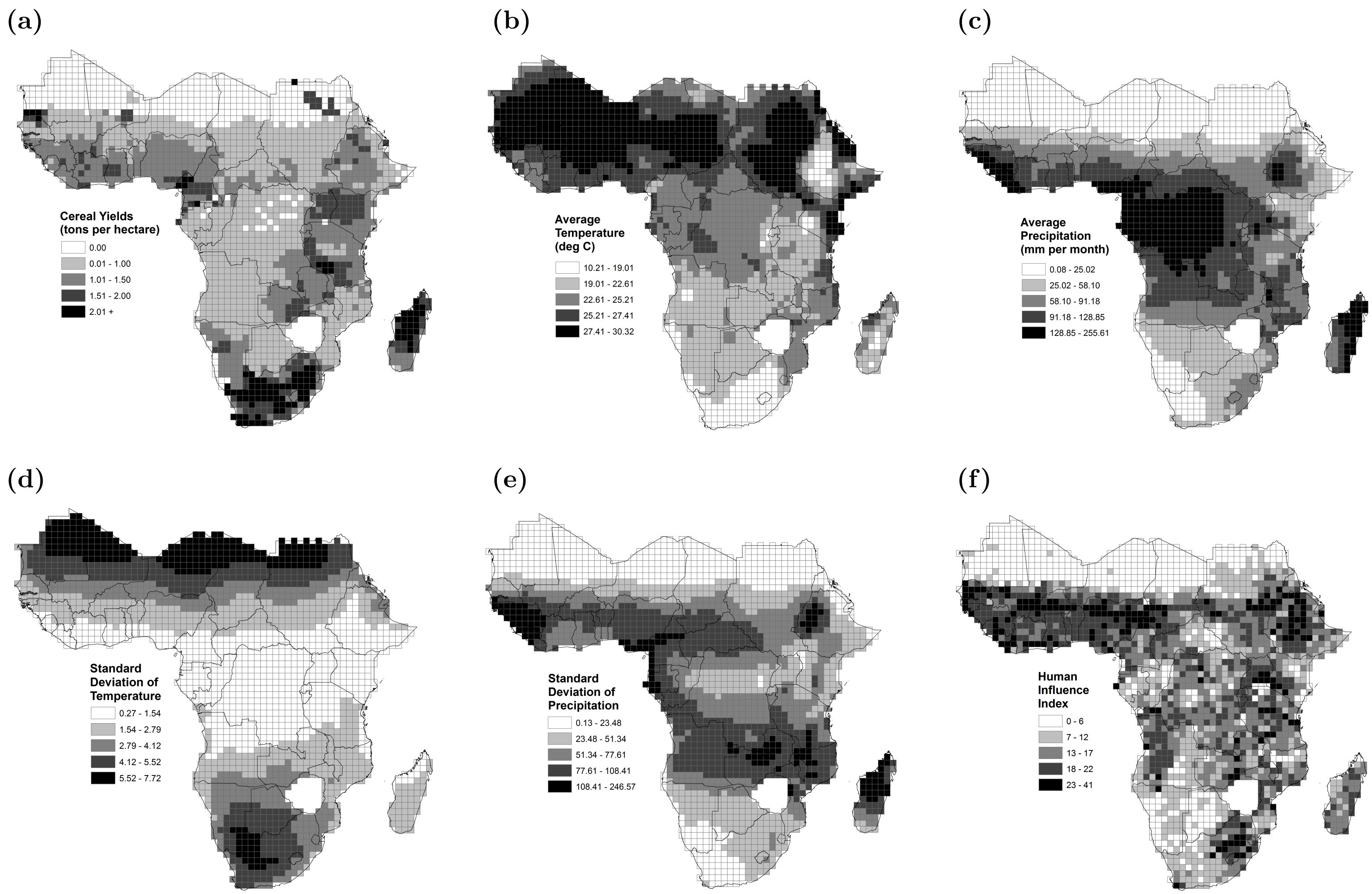

Figure 1: Distribution of Key Variables in Sub-Saharan Africa

Sources: (a) Monfreda et al. (2008); (b)-(e) Nordhaus et al. (2006); (f) CIESIN (2011)