Generating Plausible Crop Distribution Maps for Sub-Sahara Africa Using Spatial Allocation Model

Liangzhi You*
International Food Policy Research Institute
2003 K Street, NW, Washington, DC 20006
Phone: (202) 862-8168
Email: l.you@cgiar.org

Stanley Wood
International Food Policy Research Institute
2003 K Street, NW, Washington, DC 20006
Phone: (202) 862-8122
Email: s.wood@cgiar.org

Ulrike Wood-Sichra
7008 Darby Towne Court
Alexandria, VA 22315
Phone: (703) 924-0824
Email: u.wood-sichra@cgiar.org

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Denver, Colorado, August 1-4, 2004

*Corresponding author

ABSTRACT

Spatial data, which are data that include the coordinates (either by latitude/longitude or by other addressing methods) on the surface of the earth, are essential for agricultural development. As fundamental parameters for agriculture policy research agricultural production statistics by geopolitical units such as country or sub-national entities have been used in many econometric analyses. However, collecting sub-national data is quite difficult in particular for developing countries. Even with great effort and only on regional scales, enormous data gaps exist and are unlikely to be filled. On the other hand, the spatial scale of even the subnational unit is relatively large for detailed spatial analysis. To fill these spatial data gaps we proposed a spatial allocation model. Using a classic cross-entropy approach, our spatial allocation model makes plausible allocations of crop production in geopolitical units (country, or state) into individual pixels, through judicious interpretation of all accessible evidence such as production statistics, farming systems, satellite image, crop biophysical suitability, crop price, local market access and prior knowledge. The prior application of the model to Brazil shows that the spatial allocation has relative good or acceptable agreement with actual statistic data. The current paper attempts to generate crop distribution maps for Sub-Sahara Africa for the year 2000 using the spatial allocation model. We modified the original model in the following three aspects: (1) Handle partial subnational statistics; (2) Include the irrigation map as another layer of information in the model; (3) Add subsistence portion of crops in addition to the existing three input and management levels (irrigated, high-input rainfed and low-input rainfed). With the modified spatial allocation model we obtain 5 by 5 minutes resolution maps for the following 20 major crops in Sub-Sahara Africa: Barley, Beans, Cassava, Cocoa, Coffee, Cotton, Cow Peas, Groundnuts, Maize, Millet, Oil Palm, Plantain, Potato, Rice, Sorghum, Soybeans, Sugar Cane, Sweet Potato, Wheat, Yam. This approach demonstrates that remote sensing technology such as satellite imagery could be quite useful in improved understanding of the spatial variation of land cover, agricultural production, and natural resources.

JEL classification: C60; Q15; Q24

Key Words: Sub-Sahara Africa, cross entropy, satellite image, spatial allocation, agricultural production, crop suitability

1. Introduction

In the design and targeting of rural development strategies to stimulate economic growth and alleviate poverty, we have realized that it is very important to conserve natural resource base in order to maintain long-term sustainable growth. Since location matters from an agricultural perspective (as most other things too), the impact of the development strategies depends, in large extent, upon our better understanding of spatial determinants of agricultural development (Wood, Sebastian, Nachtergaele, Nielsen and Dai, 1999; Sebastian and Wood, 2000). Spatial data (sometimes referred to as geo-referenced data), which are data that include the coordinates (either by latitude/longitude or by other addressing methods) on the surface of the earth, are essential for any meaningful development strategies. More and more agricultural economists argue for the importance of spatial data and actually use spatial analysis in their research (Nelson, 2002; Staal, Baltenweck, Waithaka, deWolff and Njoroge , 2002; Luijten, 2003; Bell and Irwin, 2002; Anselin, 2002). As fundamental parameters for agriculture policy research agricultural production statistics by geopolitical units such as country or sub-national entities have been used in many econometric analyses. However, collecting sub-national data is quite difficult in particular for developing countries. Even with great effort and only on regional scales, enormous data gaps exist and are unlikely to be filled. On the other hand, the spatial scale of even the subnational unit is relatively large for detailed spatial analysis. To fill these spatial data gaps we proposed a spatial allocation model (You and Wood, 2003). Using a generalized cross-entropy approach, our spatial allocation model makes plausible allocations of crop production in geopolitical units (country, or state) into individual pixels, through judicious interpretation of all accessible evidence such as production statistics, farming systems, satellite image, crop biophysical suitability, crop price, local market access and prior knowledge. The application of the model to Brazil shows that the spatial allocation has relative good or acceptable agreement with actual statistic data.

The current paper attempts to generate crop distribution maps (spatial production data) for Sub-Sahara Africa for the year 2000 using the spatial allocation model. However, we must make some changes in the original model first. In the above Brazil case, we ran the spatial allocation for each of the 28 states of Brazil because we have complete state-level production data (yield and area) for all the crops and all the states. However, collecting sub-national data in Sub-Sahara Africa (SSA) has been quite difficult even with lots of effort and time. Since the spatial allocation model *simultaneously* allocates all crops into the pixels within the spatial allocation unit (country or state), complete production data on the level of that spatial allocation unit are needed. For SSA it is only possible at country-level. Therefore, for the current SSA production allocation, the spatial allocation unit is country. On the other hand, we do want to include the existing sub-national data, no matter how spotty they are, to guide our allocation. So we modified our original spatial allocation model to deal with partial sub-national data, which may be more realistic for most problems than our Brazilian case. In addition, we also add the irrigation map as another layer of information into our model.

In the following, we will first introduce different types of information which are included in the model. Second, we will build the spatial allocation model using cross-entropy approach as we did before (You and Wood, 2003). Third, we apply the modified model to Sub-Sahara Africa and the results will be crop distribution maps for the selected crops. Finally we conclude with some remarks on the possible application of the results and on how to further improve the model.

2. Data

The total agricultural land in SSA is about 903 million hectares in 2000, which covers about 40% of land surface and makes agriculture the most important land uses in SSA. The majority of agricultural land, about 82%, is permanent pasture while the total land for annual and permanent crops are only about 161 million hectares (FAOSTAT, 2003). SSA locates mostly in humid or sub-humid tropic zones with heterogeneous cover such as mixed forest, pasture, permanent and annual crop land. The majority of farmers are smallholders or even subsistence farmers. Some staple crops such as plantain, cassava are commonly planted in farmers' backyards. All these result in quite a challenge for lots of our data introduced in the following. In our current allocation, we choose Year 2000 as our base year, and so all time-dependent data will be based on 2000, or a three-year average from 1999 to 2001.

We selected the following 20 crops for our spatial allocation for Sub-Sahara Africa: Barley, Beans, Cassava, Cocoa, Coffee, Cotton, Cow Peas, Groundnuts, Maize, Millet, Oil Palm, Plantain, Potato, Rice, Sorghum, Soybeans, Sugar Cane, Sweet Potato, Wheat, Yam

2.1 Production statistics

The country-level production data are available from FAO. Even with our great effort, the subnational data[1] coverage of our base year (2000) is quite spotty. Figure 2.1 shows the subnational data coverage for the 20 selected crops. Just a few countries, namely Benin, Cameroon, D.R. Congo, Uganda, Zambia, Mozambique, have over half crops with available subnational data. Quite a few countries such as Angola, Republic of Congo, Gabon, Ivory Coast have only national data available. Across crops, cow peas, bean, maize and cassava,

---

[1] In this paper subnational unit refers to the first geopolitical level under country such as districts in Uganda, regions in Nigeria, provinces in South Africa. The second level sub-national data are hardly available for SSA.

which available subnational data cover over 70% of total subnational units, have more subnational data than the rest crops. Overall we only have approximately 40% coverage of complete subnational data.

[Figure 2.1 Subnational data coverage map]

2.2 Production pattern

External inputs such as irrigation, fertilizer, pesticide, affect agricultural production in many ways. Therefore, disaggregating the crop area into different production systems according to the input level could potentially improve the spatial allocation. In addition, as we will find in the following section, agroecological suitability is also defined according to different input levels. For those area statistics we have, either on country-level or on sub-national level, we partition each crop production into four levels: irrigated, rainfed–high input level, rainfed-low input level and subsistence. Although subsistence crops are almost all low-input rainfed crops in SSA, we list it separately to better allocate these crops to where people live without considering crop suitability or even potential revenue. To break crop areas into the above four production systems remains quite a challenge due to lack of reliable data. We do that by a mixture of aggregate data, informal studies, and expert opinion.

2.3 Population density

Profit relations are amongst the main determinants of the type and volume of agricultural production activities and play a fundamental role in formulation of development plans and related decisions. For traded crops, farmers' decision to grow what and where depends very much on the profitability of the crop production. Even for subsistence crops the values of crop productions are also important factor for farmers to make planting decision. Market, even as imperfect market as in Sub-Sahara Africa (Khernallah, Delgado, Gabre-

Madhin, Minot and Johnson, 2000), is important factor in the evolution of agricultural production system. Another determinant factor is market access which is an indicator of transaction cost from the location to the nearby markets (Bolwig and Wood, 2002). The market access affects both cost of production (via input market such as fertilizer, pesticide, seed etc) and gross revenue (via transportation cost and other transaction, for example). Currently we have no complete data for road network and market distribution in SSA, and so we simply use population density as an approximate to market access.

[Figure 2.2 Population density of Sub-Sahara Africa]

We use Gridded Population of the World (GPW) Version 2 which provides global estimates of population counts and population densities (persons per square kilometer) for 1990 and 1995 (CIESIN, IFPRI and WRI, 2000). National figures have been reconciled to be consistent with United Nations population estimates for those years. Figure 2.2 is the map of population density for Sub-Sahara Africa. In the current model, population density provides spatial variability for crop prices ($PriceVar_{ij}$ in the following Eq.(3.13)). Though the local crop prices are affected by many factors one of which is population density, the following simple linear interpolation is used to calculate $PriceVar_{ij}$ from population density data.

(2.3) $$PriceVar_{ij} = \frac{Pop_i - MinPop_k}{MaxPop_k - MinPop_k}$$

where $MinPop_k$ and $MaxPop_k$ are the population densities at 20% and 90% at Country K's cumulative population density distribution curve[2], Pop_i is the population density for pixel i. We realize that the price variation with population depend on specific crops. For example the

[2] In fact, $MinPop_k$ = Max($MinPop_k$, 5). If $MinPop_k$ is less than 5 people per square kilometer, there may be just forest with little agricultural land.

price of perishable products such as vegetable drops faster away from population center than non perishable products such as maize and rice. We keep crop identifier *j* at the equation in order to add crop-specific price variation with population in the future though it is not done here now.

2.4 Land cover image

Satellite-base land cover image is an important guide for the allocation model. As more and better remotely-sensed data become available with the new technology and the improvement of our ability to interpretate the image data, the land cover data will dramatically improve the accuracy of our production allocation. The time when satellite takes these land use data should coincide with the chosen year of our allocation. For Sub-Sahara Africa, we use Africa Land Cover 2000 from Global Land Cover 2000 project (GLC2000). GLC2000 makes use of VEGA 2000 data: a dataset of 14 months (November 1 1999 – December 31 2000) of pre-processed daily global data acquired by the VEGETATION instrument on board SPOT 4, made available through a sponsorship from members of the **VEGETATION** programme (http://vegetation.cnes.fr/).

[Figure 2.3 Agricultural extent for Sub-Sahara Africa]

2.5 Global irrigation maps

The Land and Water Development Division of Food and Agriculture Organization of the United Nations and the Center for Environmental Systems Research of the University of Kassel, Germany, have been co-operating in the development of a global irrigation mapping facility. The final result is a global irrigation map, which shows the amount of area equipped for irrigation around 1995 as a percentage of the total area on a raster with a resolution of 5 minutes. The way the map is generated is not uniform across countries. Due to the map generation method, the quality of the map can never be uniform. The overall quality of the map depends heavily on the individual quality of the data for the different countries. The area actually irrigated in 1995 was smaller, but is unknown for most countries.

[Figure 2.4 Irrigation map for Sub-Sahara Africa]

In the current spatial allocation, we use the irrigation map as another layer to inform the model where to allocate the irrigated areas. In Sub-Sahara Africa, rainfed agriculture dominates, and irrigation is quite spare as shown in Figure 2.4 (d). South Africa and Madagascar are the two countries with major irrigated areas, about 1.43 million hectares and 1.15 million hectares respectively. Countries such as Mauritania, Chad, Congo D.R. and Namibia have little irrigation.

2.6 Agroclimatic crop suitability

Different crops have different thermal, moisture and soil requirements, particularly under rainfed conditions. The Food and Agriculture Organization of the United Nations (FAO) with the collaboration of the International Institute for Applied Systems Analysis (IIASA), has developed the Agro-ecological Zones (AEZ) methodology that enables rational land-use planning on the basis of an inventory of land resources and evaluation of biophysical limitations and potentials. AEZ methodology provides a standardized framework for the characterization of climate, soil and terrain conditions relevant to agricultural production. Crop modeling and environmental matching procedures are used to identify crop-specific limitations of prevailing climate, soil and terrain resources, under assumed levels of inputs and management conditions. The AEZ methodology provides maximum potential and agronomically attainable crop yields and suitable crop areas for basic land resources units (usually grid-cells in the recent digital databases). (Fischer et al 2001; FAO, 2003). We only use the following three production system types in FAO/IIASA suitability datasets: Irrigated – high input (we simply call it "irrigated"), Rainfed – high input, Rainfed – low input. For each crop by the above three input levels, we define our suitable land as the sum of the following four classes: very suitable, suitable, moderate suitable and marginal suitable. Correspondingly, the yield is calculated as the area-weighted average of the above four suitable classes. (FAO, 1981; FAO 2003). Some crops have many types, such as highland and lowland maize germplasm, sub-divided by maturity class. The single "maize" crop surface is a composite in which each pixel would use the best variety. Figure 2.3 shows the suitability surfaces for maize in SSA. We show both potential yield and suitable area distributions under either high input or low input rainfed conditions. As we can see, maize is widely suitable in Sub-Sahara African except in the extremely northern countries such as Mauritania, Mali, Niger, Chad and extremely southern countries such as Angola, Botswana and South Africa. Maize normally doesn't require irrigation, which is shown by the less suitable areas for irrigated maize (Figure 2.4 (c)). However, the irrigated maize has a much higher potential yield than rainfed maize. The average potential yield for irrigated maize in Sub-Sahara Africa could be as high as 390kg/ha while the corresponding figures for high-input and low-input rainfed maize are 290kg/ha and 40kg/ha (Figure 2.6).

[Figure 2.5 Suitability surfaces of maize - Area]

[Figure 2.6 Suitability surfaces of maize - Yield]

3. Modified Spatial Allocation Model

3.1 Cross-entropy approach

The origin of entropy goes back to famous Boltzmann's distribution law in thermodynamics (Jaynes, 1979). It is a measure of "disorder" of molecules in a system. One of the fundamental law of nature is the second law of thermodynamics, which says the entropy of a closed system never decreases and increases whenever possible. Shannon (1948) introduced information entropy to measure the uncertainty (state of knowledge) of the expected information, which gives birth to the science later called information theory. In information theory, information is not what you know but uncertainty, that is, information is statistical property of a message. Any probability distribution p_i, $i = 1, 2, \ldots, n$, of a random variable provides some *information* about that variable. Intuitively, information should be a decreasing function of p_i because the more unlikely an event, the more interesting it is to know that it can happen (p.34-35, Sen, 1975; Bera and Billias, 2002). Shannon defined entropy $H(p)$ as a weighted sum of the information $-\ln p_i$, $i = 1,2,\ldots,n$ with respective probabilities as weights:

$$H(p) = -\sum_{i=1}^{n} p_i \ln p_i = -E(\ln p) \tag{3.1}$$

with convention that $0 \ln 0 = 0$. $E(\ln p)$ is expected value of $\ln p$.

Jaynes (1957) adopted information entropy concept and proposed the maximum entropy principle in statistical inference: the least informative probability distribution p_i can be found by maximizing the entropy $H(p)$. In (3.1), the solutions are: $p_i = 1/n$, $i = 1, 2, \ldots, n$, $H(p) = \ln n$. Put the principle in other words: in the absence of information to the contrary, all possible states of system are equally likely. Generalized maximum entropy (GME) approach is based upon this principle.

Following (3.1), the cross-entropy of one probability distribution $p=\{p_1, p_2, \ldots, p_n\}$with respect to another probability distribution $q=\{q_1, q_2, \ldots, q_n\}$ can be defined

(3.2) $$CE(p,q) = -\sum_{i=1}^{n} p_i \ln p_i / q_i = E(\ln p) - E(\ln q)$$

This is actually a measure of distance between two probability distributions p and q. If we choose the non-informative q, i.e, $q = \{1/n, 1/n, \ldots, 1/n\}$, then $CE(p, q)$ becomes:

(3.3) $$CE(p,q) = \sum_{i=1}^{n} p_i \ln p_i - \ln n = -H(p) - \ln n$$

Therefore maximizing entropy is in fact a special case of minimizing cross-entropy with respect to a uniform distribution. The cross entropy (CE) approach can be stated as a minimization problem where the objective function is the cross entropy and the constraints are some side conditions, the prior knowledge.

A comprehensive book by Golan, Judge and Miller (1996) has brought much interest in applying the entropy approach (Lencer and Miller, 1998; Paris and Howlitt, 1998; Robinson, Cattaneo and El-Said, 2000; Zhang and Fan, 2001). The unique feature of the entropy approach is to overcome two empirical problems that hamper traditional econometrics: multi-collinearity and ill-posed problems (underdetermined or partial incomplete data) (Paris and Caputo, 2001; Golan, Judge and Miller, 1996). The idea is to remove irrelevant information at the beginning of a problem rather than taking pains to make dubious assumptions. Preckel (2001) compares least squares and entropy methods from a penalty function perspective, and concludes that the differences between these two approaches boil down to how the supports for errors and coefficients are defined in generalized cross-entropy approach. When the supports are specified to be symmetric, wide, and centered on zero for the residual errors, the coefficient estimates are essentially indistinguishable (Preckel, 2001). Shen and Perloff (2001) estimate a ratio of parameters using different methods and concludes that GME (and Bayesian method of moments) estimator has a much smaller mean square errors and average biases than do ordinary least squares (OLS). A recent article by Bera and Bilias (2002) does an excellent synthesis on different estimation approaches such as method of moments, maximum entropy, maximum likelihood, empirical likelihood, estimating function and generalized methods of moments. The article unifies many of these estimation techniques in the framework of Cressie and Read (1984) power divergence criterion, and puts these techniques in an interesting historical perspective.

3.2 Modified spatial allocation model

We built our spatial allocation model on the cross entropy approach (You and Wood, 2003). Now we modified the model to accommodate partial sub-national data and new irrigation maps. Let s_{ijl} be the area share allocated to pixel i and crop j at input leve l with a certain country (say **X**) in SSA. $CropArea_{jl}$ is the total physical area for crop j at input level l, and A_{ijl} the area allocated to pixel i for crop j at input level l in country **X.** Therefore:

(3.4) $$s_{ijl} = \frac{A_{ijl}}{CropArea_{jl}}$$

Let π_{ijl} be the prior area shares we know by our best guess for pixel i and crop j at input level l in country **X**. The modified spatial allocation model can be written as follows:

(3.5) $$\underset{\{s_{ijl}\}}{MIN} \quad CE(s_{ijl}, \pi_{ijl}) = \sum_i \sum_j \sum_l s_{ijl} \ln s_{ijl} - \sum_i \sum_j \sum_l s_{ijl} \ln \pi_{ijl}$$

subject to:

(3.6) $$\sum_i s_{ijl} = 1 \quad \forall j \, \forall l$$

(3.7) $$\sum_j \sum_l CropArea_{jl} \times s_{ijl} \le Avail_i \quad \forall i$$

(3.8) $$CropArea_{jl} \times s_{ijl} \le Suitable_{ijl} \quad \forall i \, \forall j \forall l$$

(3.9) $$\sum_{i \in k}\sum_{l} CropArea_{jl} \times s_{ijl} = SubCropArea_{jk} \quad \forall k \forall j \in J$$

(3.10) $$\sum_{l \in L} CropArea_{jl} \times s_{ijl} \le IRRArea_{i} \quad \forall i$$

(3.11) $$1 \ge s_{ijl} \ge 0 \quad \forall i, j, l$$

where:

i : *i = 1, 2, 3, ...,* pixel identifier within the allocation unit, and

j: *j = 1, 2, 3, ...,* crop identifier (such as maize, cassava, rice) within the allocation unit, and

l: l = irrigated, rainfed-high input, rainfed-low input, subsistence, management and input levels for crops

k: k = 1, 2, 3, …, identifiers for sub-national geopolitical units

J: a set of those commodities which sub-national production statistics exist

L: a set of those commodities which are partly irrigated within pixel *i*.

$Avail_i$: total agricultural land in pixel *i*, which is equal to total agricultural area estimated from land cover satellite image as described in the previous section.

$Suitable_{ijl}$: the suitable area for crop *j* at input level *l* in pixel *i*, which comes form FAO/IIASA suitability surfaces as introduced in the previous section.

$IRRArea_i$: the irrigation area in pixel *i* from global map of irrigation (Siebert et al, 2001).

Comparing to the original spatial allocation model (You and Wood, 2003), there are two new constraints: equations (3.9) and (3.10). Constraint (3.9) sets the sum of all allocated areas within those subnational units with existing statistical data to be equal to the corresponding subnational statistics. Constraint (3.10) includes the irrigation information: the sum of all allocated irrigated areas in any pixel must not exceed the area equipped for irrigation indicated in global map of irrigation (Siebert et al, 2001). The objective function and all other equations are similar to the original model. This modified model could deal with incomplete data coverage at subnational levels, and in this sense it broaden the scope of spatial allocation model. Incomplete data may be the rule rather than the exception in real world, in particular in developing countries.

Obviously a informed prior(π_{ijl}) is very important for the success of the model. We create the prior based upon the available evidence. First for each pixel, we calculate the potential revenue as

$$(3.12) \qquad Rev_{ijl} = Price_j \times Pricevar_{ij} \times Yield_{jl} \times Suitability_{ijl} \times Suitable_{ijl}$$

where $Price_j$ and $Yield_{jl}$ are the price index and the average yield for crop j at input level l (yield only) for the allocation unit (countries in SSA), $Suitability_{ijl}$ is the suitability for crop j at input level l and pixel i, which is represented as proportion (value between 0 and 1) of the optimal yield as described in Section 2.7. $Pricevar_{ij}$ is the price variability (value between 0 and 1) for crop j and pixel i. Currently we use the population density as an approximate to spatial price variation. Then we pre-allocate the available statistical crop areas (at various geopolitical scales) into pixel-level areas by simple weighting:

$$(3.13) \qquad Area_{ijl} = SubCropArea_{jk} \times Percent_{jl} \times \frac{Rev_{ijl}}{\sum_{i \in k} Rev_{ijl}} \quad \forall j \forall i \forall l$$

where $Areai_{jl}$ is the area pre-allocated to pixel i for crop j at level l, $Percent_{jl}$ is the area percentage of crop j at input level l. For those geopolitical units without area statistics, we simply merge them together and obtain the total area for that merged unit by subtracting the sum of available subnational areas from national total. After this pre-allocation, we calculate the prior by normalizing the allocated areas over the whole country.

(3.13)
$$\pi_{ijl} = \frac{Area_{ijl}}{\sum_i Area_{ijl}} \quad \forall j \forall i \forall l$$

4. Results

The application of the spatial allocation model faces two major challenges. First one is the inconsistency among the various constraints due to imperfect data. For example, the sum of all the statistical crop areas may be even larger than the cropland available from satellite image at either national or subnational levels. Therefore, the constraint (3.7) in the model are directly conflicting to constraint (3.6) if at national level or constraints (3.9) if at subnational level, and the optimization problem becomes infeasible. Similar conflicts may appear among the rest of the constraints: less available irrigated area than the sum of statistical irrigated crops areas ((3.6) or (3.9) vs. (3.10)), less cropland than the irrigated area ((3.7) vs. (3.10)), less suitable areas than the statistical crop areas ((3.8) vs. (3.6) or (3.9)). These conflicting constraints have to be solved before we could run the model. Actually, these inconsistencies happen for every SSA country in the current application. We wrote computer programs (in FoxPro language) to deal with these conflicts. In principle, we treat the production/area statistics as the "truth"[3], and then modify the areas from other sources. For example, we scale

[3] We don't claim that the statistics (either national or subnational) are more reliable than say, satellite image. We set statistics as a benchmark to make the results more comparable since these statistics are widely used and recognized.

up cropland areas if they are less than the sum of statistical crop areas. For those pixels where there is zero cropland but positive irrigated areas, we set the cropland areas equal to the irrigated areas.

The second challenge is the huge size of the optimization problem, in particular for big countries. With the current grid resolution of five minutes, middle-size country such as South Africa has about 10,000 pixels with nonzero agricultural land (pixels with zero agricultural land are excluded from the model run). South Africa produces 16 of the 20 crops considered, and each crop has four input levels. Therefore the total number of unknowns for South Africa is in the range of 600,000, which is relatively large for a nonlinear programming problem. Large countries such as Angola and Congo, D.R. are even larger problems. In addition the objective function with logarithms is a challenge for any nonlinear programming solver. GAMS (2003) is used to solve the model. The requirement for CPU time depends on specific countries, ranging from a few minutes to over 50 hours on our Dell desktop with 3.2 GHz CPU and 1GB RAM.

We run the modified spatial allocation model country by country for all 51 countries[4] in Sub-Sahara Africa. A post-processing program would take the results from GAMS and calculate both the harvest areas and productions by pixels. Figure 3.1 shows the crop area distribution maps for all the 20 crops considered. These are the 5 x5 minutes (about 9000 hectares in Africa) crop distribution maps.

[Figure 4.1 Estimated crop distribution maps of Sub-Sahara Africa]

5. Final Remarks

[4] Some island countries such as Mayotte, Seychelles have no or little agriculture productions.

We have proposed a spatial allocation model of crop production based on a cross-entropy approach (CE). The approach utilizes information from various sources such as best available production statistics, satellite imagery, biophysical crop suitability assessments, irrigation map, as well as population density, in order to generate plausible, disaggregated estimates of the distribution of crop production on a pixel basis. With this spatial allocation model we obtain 5 by 5 minutes resolution maps for the 20 major crops in Sub-Sahara Africa. We also find that new technologies such as remote sensing and image processing prove to be useful tools for exploring the spatial heterogeneity of agriculture production, infrastructure and natural resources. On the other hand, working at a spatial scale of individual pixels creates many data management and computational challenges. Some of these challenges need to be met through improved numerical methods and mathematical optimization software.

Though the current model provides what appear, in the absence of "truth" regarding the real distribution of production, to be reasonable results, more work is underway to improve its performance. The obvious way forward is to improve the underlying quality of the parameters currently included in the model, since the end results can only be as accurate as the input information. These include better approximations of the agricultural extent, more realistic crop suitability surfaces, and more research on the association between crop production and population density. On the other hand, we could also add more information into the model. For example, household or agricultural survey information on the location and quantity of crop production would provide a direct, sampled calibration of the entire crop distribution surface. If such information exists and it is of reasonable quality, it will definitely improve the estimation accuracy. We could also add some other behavioral assumptions. For example, it seems reasonable to assume that farmers would opt to plant a higher revenue crops in any given location, all other things being equal. But potential revenue is in reality a proxy for potential profitability, and some could argue that risk minimization might also play a role. Thus there are several options for further work in exploring alternative drivers of crop choice, both individually and in crop combinations, in each location.

Figure 2.1 Subnational data coverage

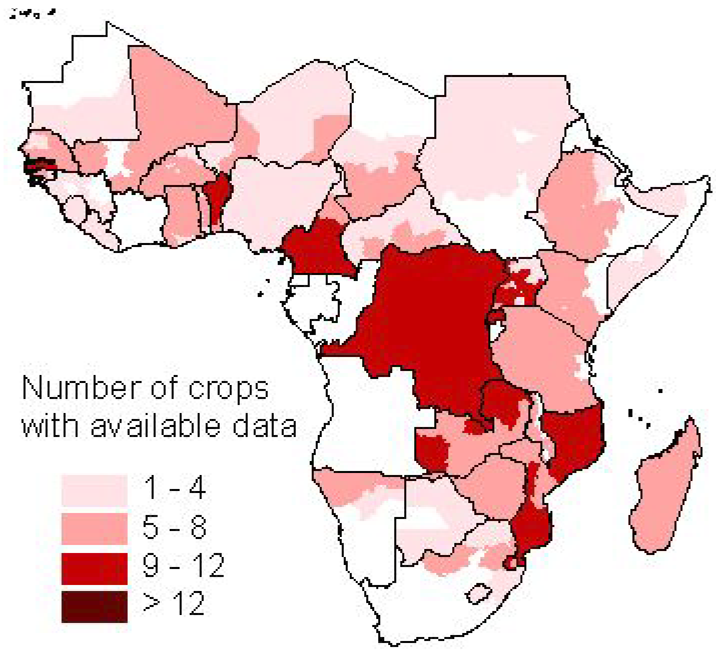


Figure 2.2 Population density for Sub-Sahara Africa

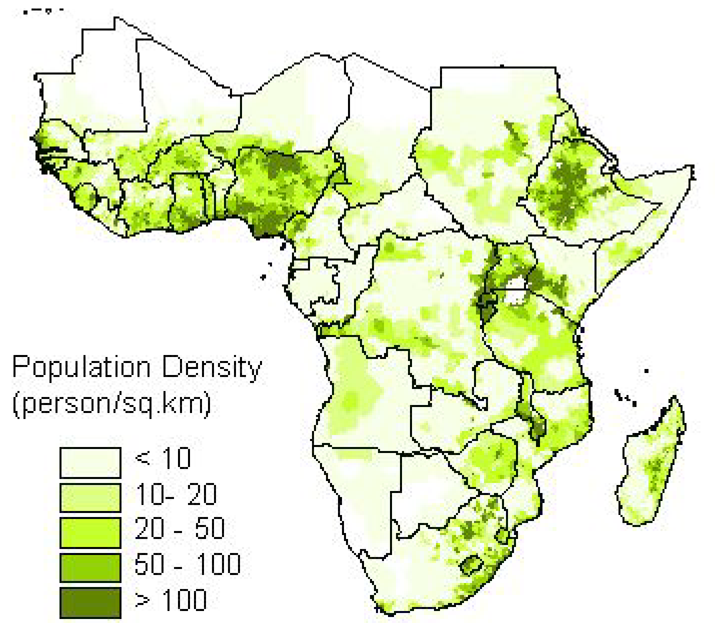


Figure 2.3 Agricultural land

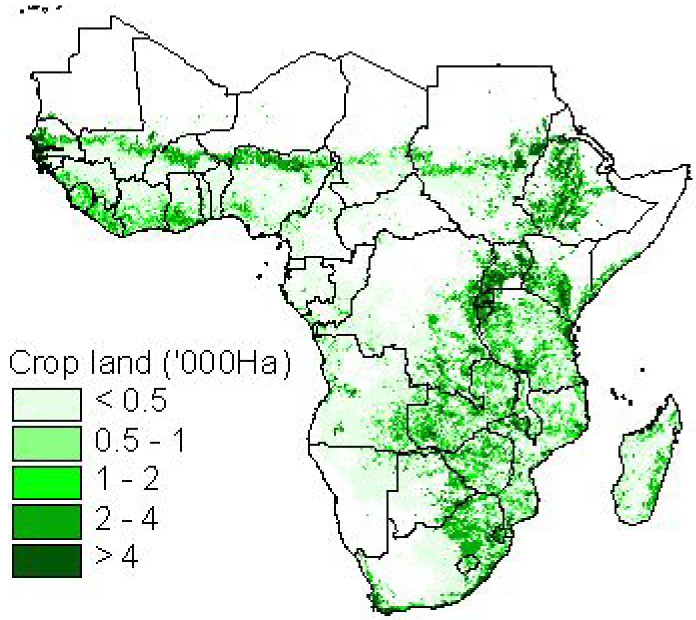


Figure 2.4 Irrigation map

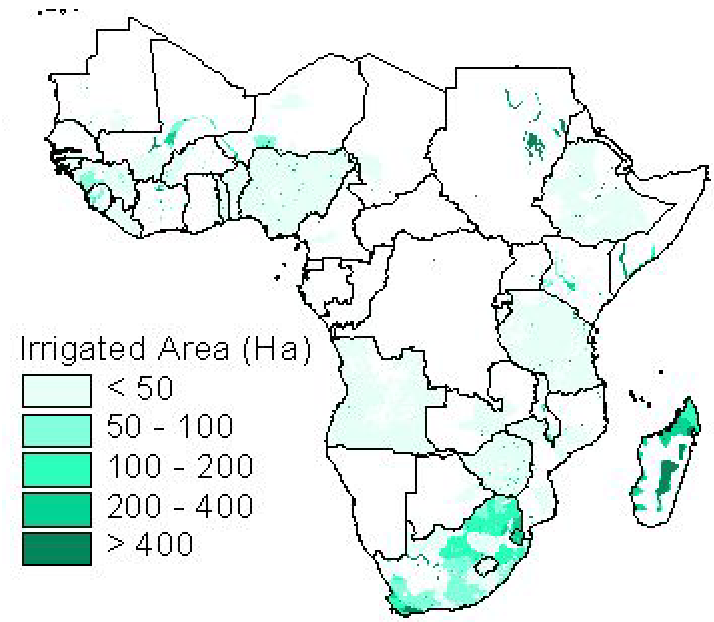


Figure 2.5 Crop suitability surfaces – suitable areas

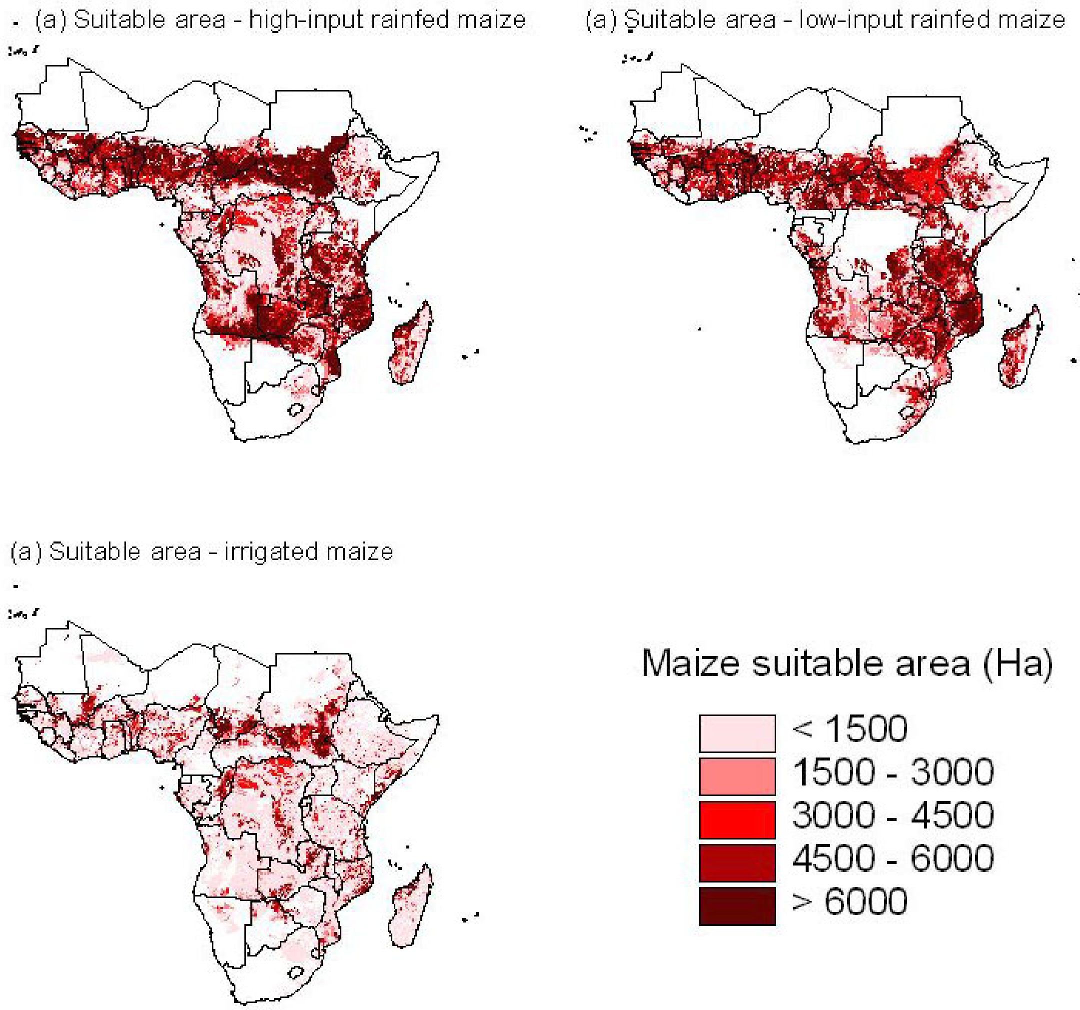


Figure 2.6 Crop suitability surfaces – potential yield

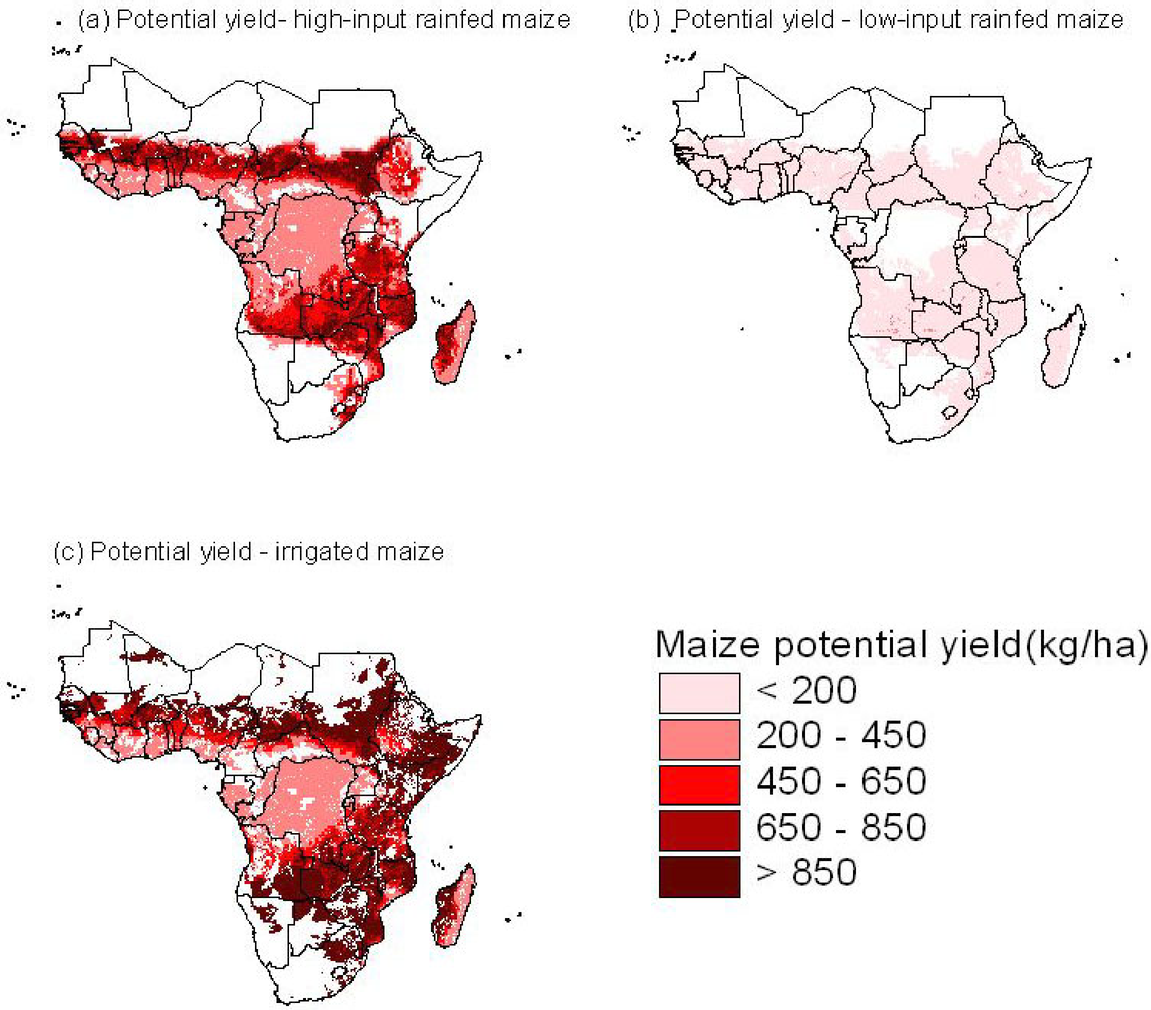
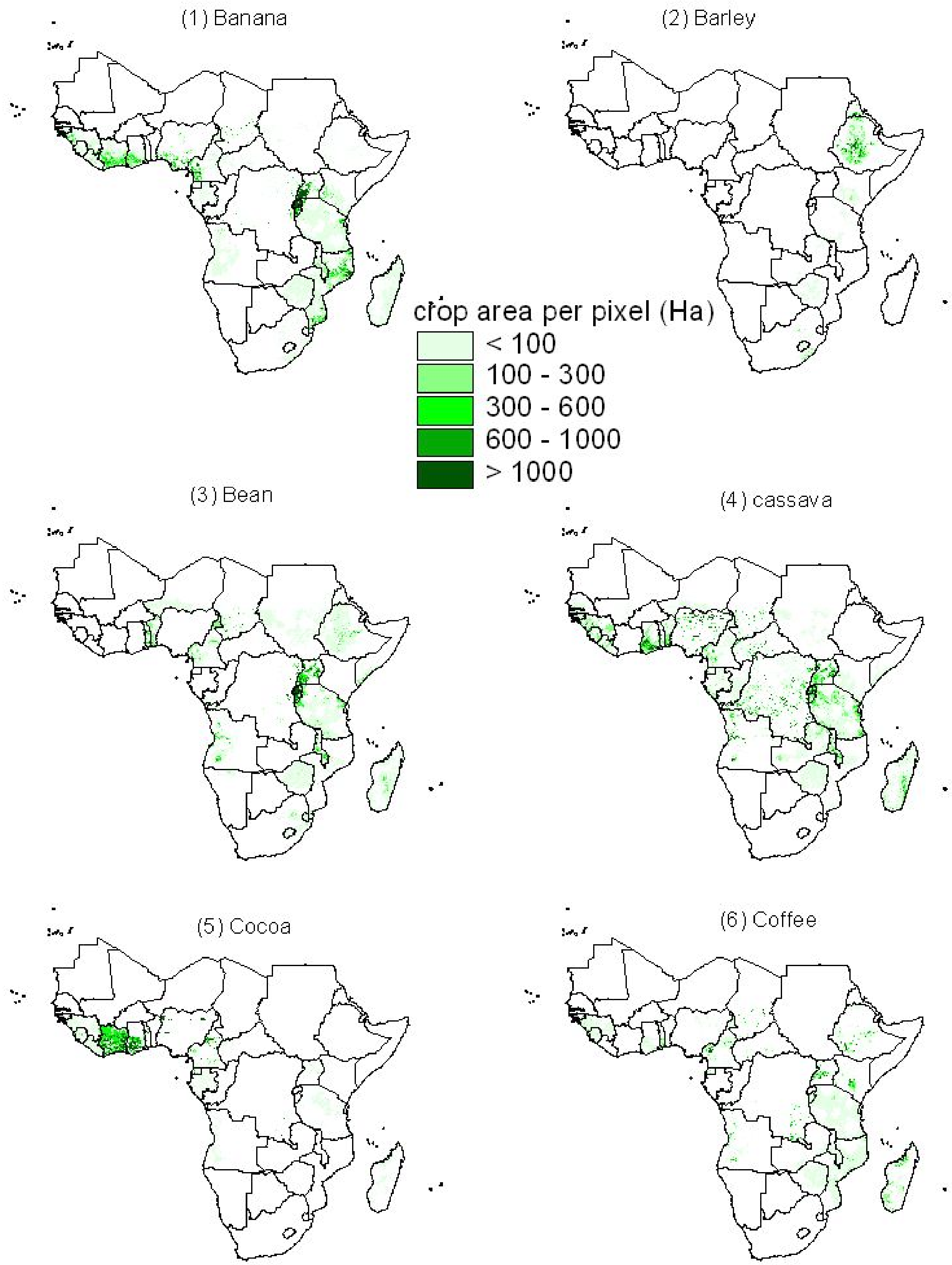


Figure 3.1 Estimated crop distribution maps of Sub-Sahara Africa



REFERENCES

Anselin, L., 2002. Under the hood: issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27

Bell, K.,Irwin, E.G., 2002. Spatially explicit micro-level modeling of land use change at the rural-urban interface. *Agricultural Economics 27*

Bera, Anil K. and Yannis Bilias, 2002. The MM, ME, ML, El, EF and GMM approaches to estimations: a synthesis. *Journal of Econometrics*, 107, p.51-86

Center for International Earth Science Information Network (CIESIN), Columbia University; International Food Policy Research Institute (IFPRI); and World Resources Institute (WRI). 2000. Gridded Population of the World (GPW), Version 2. Palisades, NY: CIESIN, Columbia University. Available at http://sedac.ciesin.columbia.edu/plue/gpw.

Cressie, N. and T.R.C. Read, 1984. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society*, Series B 46, p.440-464

Fischer, Gunther, M. Shah, H. Velthuizen, F. Nachtergaele, 2001. *Global Agro-ecological Assessment for Agriculture in the 21st Century*, International Institute for Applied Systems Analysis, Laxenburg, Austria

FAO, 1981. *Report of the Agro-Ecological Zones Project*, World Soil Resources Report No 48, Vol.1-4, Rome, FAO

FAO, 2003. http://www.fao.org/ag/AGL/agll/gaez/index.htm

GAMS, 2003. http://www.gams.com/.

Golan, Amos, G. Judge and D. Miller, 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, New York: John Wiley & Sons

Jaynes, E. T., 1957. *Information Theory and Statistical Methods I*, Physics Review, 106(1957): 620-630

Jaynes, E. T., 1979. Where do we Stand on Maximum Entropy?. in *The Maximum Entropy Formalism*, R. D. Levine and M. Tribus (eds.), MIT Press, Cambridge, MA, p. 15

Keherallah, M., C. Delgado, E. Gabre-Madhin, N. Minot, and M. Johnson, 2000. The Road Half Traveled: Agricultural Market Reform in Sub-Sahara Africa. Food Policy Report, International Food Policy Research Institute, Washington, DC, USA.

Lence, S. H. and D. J. Miller, 1998. Estimation of multi-output production functions with incomplete data: a generalised maximum entropy approach. *European Review of Agricultural Economics*, 25, 188-209

Luijten, J.C., 2002. A systematic method for generating land use patterns using stochastic rules and basic landscape characteristics: results from a Colombian hillside watershed. *Agriculture Ecosystem & Environment*, 95, p. 427-441

Nelson, G.C. 2002. Introduction to the special issue on spatial analysis for agricultural economists, *Agricultural Economics* 27, p.197-200

Paris, Q., R.E. Howitt, 1998. Analysis of ill-posed production problems using maximum entropy. *American Journal of Agricultural Economics*, 80, p.124-138

Preckel, Paul V., 2001. Least squares and entropy: a penalty function perspective. *American Journal of Agricultural Economics*, 83 (2), p.366-377

Robinson, Sherman, A. Cattaneo, and E. Moataz, 2000. *Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods.* Trade and Macroeconomics Division Dicussion Paper No.58, International Food Policy Research Institute, Washington D.C.

Sebastian, Kate and S. Wood, 2000. *Spatial Aspects of Evaluating Technical Change in Agriculture in Latin America and the Caribbean*. Research report, International Food Policy Research Institute, Washington, D.C.

Shen, Edward. and J. M. Perloff, 2001. Maximum entropy and Bayesian approaches to the ratio problem. Working paper January 2001, Department of Agricultural and Resource Economics, University of California Berkeley.

Staal, S.J., I. Baltenweck, M.W. Waithaka, T. deWolff, L. Njoroge, 2002. Location and uptake: intergrated household and GIS analysis of technology adoption and land use, with application to smallholder dairy farms in Kenya. *Agricultural Economics* 27, p. 295-315

Siebert, Stefan, P. Döll and J. Hoogeveen, 2001. *Global map of irrigated areas version 2.0*. Center for Environmental Systems Research, Univerity of Kassel, Germany / Food and Agriculture Organization of the United Nations, Rome, Italy

Wood, Stanley, K. Sebastian, F. Nachtergaele, D. Nielsen, and A. Dai, 1999. Spatial Aspects of The Design and Targeting of Agricultural Development Strategies, *Environment and Production Technology Division Discussion Paper No. 44,* International Food Policy Research Institute, Washington D.C.

Wood, Stanley, K. Sebastian and S. Scherr, 2000. *Pilot Analysis of Global Ecosystems: Agroecosystem*, a joint study by International Food Policy Research Institute and World Resource Institute, Washington D.C.

You, Liangzhi and S. Wood, 2003. Spatial allocation of agricultural production using a cross-entropy approach. *Environment and Production Technology Division Discussion Paper No. ??*, International Food Policy Research Institute, Washington D.C.

Zhang, X. and S. Fan, 2001. Estimating crop-specific production technologies in Chinese agriculture: a generalized maximum entropy approach. *American Journal of Agricultural Economics*, 83(2), 378-388