Generating Global Crop Distribution Maps: From Census to Grid

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

Contributed paper prepared for presentation at the International Association of Agricultural Economists Conference, Gold Coast, Australia, August 11-18, 2006

*Corresponding author

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). 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. 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 (You and Wood, 2006). The current paper attempts to generate global crop distribution maps (spatial production data) for the year 2000 using the spatial allocation 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. Third, we apply the modified model to the globe 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. Input Data Layers to Spatial Allocation Model

2.1. Crop production statistics.

While crop production data[1] at the country level are reported by Food and Agriculture Organization of United Nations (FAO), similar data within country boundaries are hardly available on a global scale. In early 2002, FAO, IFPRI and SAGE (Center for Sustainability and the Global Environment, University of Wisconsin-Madison) set up an informal collaborative consortium titled Agro-MAPS (Mapping of Agricultural Production Systems). The goal of Agro-MAPS is to compile a consistent global spatial database based upon selected sub-national agricultural statistics. Agro-MAPS is the major data source for the global sub-national crop production data in our spatial allocation, though we made a great effort to add more sub-national data. We choose Year 2000 as our base year. All time-

[1] We take a broader definition of production data. Crop production data refer to harvested area, production, and yield of a certain crop. Crop yield is defined as production divided by harvested area.

dependent input data (such as harvested area, yield) are based on 2000, or a three-year average from 1999-2001 to avoid atypical year if data available for these three years.

Considering both the global and regional importance, we selected the following 20 crops for our spatial allocation: wheat, rice, maize, barley, millet, sorghum, potato, sweet potato, cassava and yams, plantain and banana, soybean, dry beans, other pulse, sugar cane, sugar beets, coffee, cotton, other fibres, groundnuts, and other oil crops. All together, these 20 crops account for almost 90% world total crop harvested area. An aggregated crop titled "other crops" at the country level is introduced to account for all other crops beyond the above twenty ones. It is calculated by submitting the sum of all the 20 crop areas from the total arable land reported in FAOSTAT(2004).

2.2. Crop Production System

External inputs such as irrigation, fertilizer, pesticide, affect agricultural production in many ways. For example, large-scale commercial production using hybrid seeds, fertilizer, and mechanized production method generally have a much higher crop yield than the subsistence farmers who rely on traditional seed and no high inputs. Therefore, disaggregating the crop area into different production systems according to the input level could potentially improve the spatial allocation, in particular in converting the allocated area into crop production. For those area statistics we have, either on country-level or on sub-national level, we partition those crop areas into four levels according to farming technology and crop management: irrigated, rainfed–high input level, rainfed-low input level and subsistence. In the model, we include crop area shares in the above mentioned four levels.

A distinguished feature of crop production system is the hugely different yields obtained under different production systems. The observed yields reported for administrative regions are the average yields across different production systems in the region. To further disaggregate the observed yield into yields at the input levels, we collect two more indicators by crop by country: the ratio of crop yield under irrigated condition to that under rainfed condition, and ratio of yield under high-input rainfed condition to that under low input rainfed condition. These ratios vary by both countries and crop types. With these two new ratios, we could calculate the actual yields under the four input levels described in the above section.

2.3 Landcover Image

Satellite-base land cover imagery provides the most detailed spatial information on agricultural land. We will only allocate crop production within the extent of cropland. For the current allocation, we choose Global Land Cover 2000 database to estimate the crop land extent as shown in Figure 1.

[Figure 1 Global agricultural extent]

2.4. Agroclimatic Crop Suitability

While the land allocation to different crop productions is to a large extent determined by demographic, socioeconomic, cultural, and political factors, the range of crop land uses for growing certain crops is also limited by environmental factors including climate, topography and soil characteristics. The characterization of these conditions could provide helpful guidance on the location of the crop growing areas within the administrative regions. We tend to the agroclimatic crop suitability surfaces from global agro-ecological zones (AEZ) study. In the recent study (Fischer et al 2001; FAO 2003), the AEZ methodology provides maximum potential and agronomically attainable crop yields and suitable crop areas in 5 by 5 minutes grid-cells.

2.5 Population density

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. We use population density as a proxy to market access for the crop allocation, and for subsistence portions of the crops, population density would directly serve as the prior.

2.6 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 global irrigation map 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. In the current spatial allocation, we use the irrigation map as another layer to inform the model where to allocate the irrigated areas.

3. Spatial Allocation Model

The concept of entropy is closely related to the uncertainty embedded in a probabilistic distribution. Shannon (1948) defined entropy *H(p)* as a weighted sum of the information $-\ln p_i$, $i = 1, 2, \ldots, n$ with respective probabilities as weights:

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

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

Following (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

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

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 and the prior knowledge.

Here we define our spatial crop allocation problem in a cross entropy framework. The first thing to do is to transform all real-value parameters into a corresponding probability form. We first need to convert the reported harvested area, $HarvestedArea_{jl}$ for each crop j at input level l into an equivalent physically cropped area, $CropArea_{jl}$, using cropping intensity.

(3) $$CropArea_{jl} = HarvestArea_{jl} / CroppingIntensity_{jl}$$

Let s_{ijl} be the area share allocated to pixel i and crop j at input leve l with a certain country (say **X**). A_{ijl} is the area allocated to pixel i for crop j at input level l in country **X.** Therefore:

(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:

(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:

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

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

(8) $$CropArea_{jl} \times s_{ijl} \leq Suitable_{ijl} \quad \forall i \forall j \forall l$$

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

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

(11) $$1 \geq s_{ijl} \geq 0 \quad \forall i, j, l$$

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

The objective function of the spatial allocation model is the cross entropy of area shares and their prior. Equation (6) is adding-up constraints for crop-specific areas. Equation (7) is land cover image constraint that the actual agricultural area in pixel *i* from satellite image is the upper limit for the area to be allocated to all crops. Equation (8) is the constraint that the allocated crop area cannot exceed what are suitable for the particular crop. Constraint (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 (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 last equation, Equation (11) is basically the natural constraint of s_{ijl} as shares of total crop areas.

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

$$\text{(12)} \qquad \mathrm{Re}\,v_{ijl} = \mathrm{Pr}\,ice_j \times \mathrm{Pr}\,ice\,\mathrm{var}_{ij} \times Yield_{jl} \times Suitability_{ijl} \times Suitable_{ijl}$$

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. $Pricevar_{ij}$ 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:

$$\text{(13)} \qquad Area_{ijl} = SubCropArea_{jk} \times Percent_{jl} \times \frac{\mathrm{Re}\,v_{ijl}}{\sum_{i \in k} \mathrm{Re}\,v_{ijl}} \quad \forall j \forall i \forall l$$

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.

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

4. Results

We run the modified spatial allocation model country by country. A post-processing program would take the results from the model and calculate both the harvest areas and productions by pixels. Figure 2 shows the crop distribution maps for cereal crops among the 20 crops considered. These are the 5x5 minutes (about 9x9 km^2 on the equator) crop distribution maps. In addition to these area distribution maps, the model results include production and harvested area distribution maps as well the sub-crop type maps split by production input levels (irrigated, high-input rainfed, low-input rainfed and subsistence).

[Figure 2 Estimated crop distribution maps of the world]

5. Final Remarks

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 the world. 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. Most importantly, we have initiated with other CGIAR centers (CIAT, CYMMIT, ILRI, ICRASAT, IRRI) a systematic validation process, taking advantage of extensive field presence of CGIAR centers. The feedback from crop scientists and local experts could considerably increase the accuracy of our crop distribution maps.

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*

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.

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, 2004. http://www.fao.org/ag/AGL/agll/gaez/index.htm

FAO. 2003. *World agriculture towards 2015/2030: An FAO perspective*. FAO and EarthScan.

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

Shannon, C.. 1948. *A Mathematical Theory of Communication*, Bell System Technology Journal, 27(1948): 379-423

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.

You, L. and S. Wood. 2006. An entropy approach to spatial disaggregation of agricultural production. *Agricultural Systems*. forthcoming.

Figure 1 Agricultural land

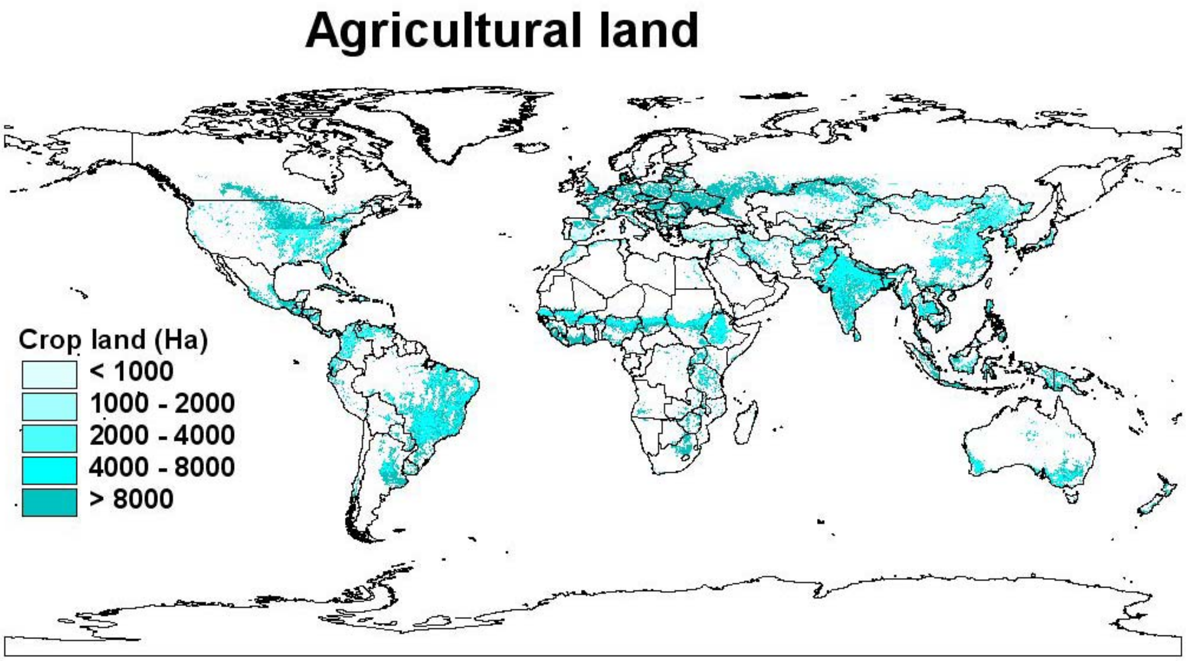
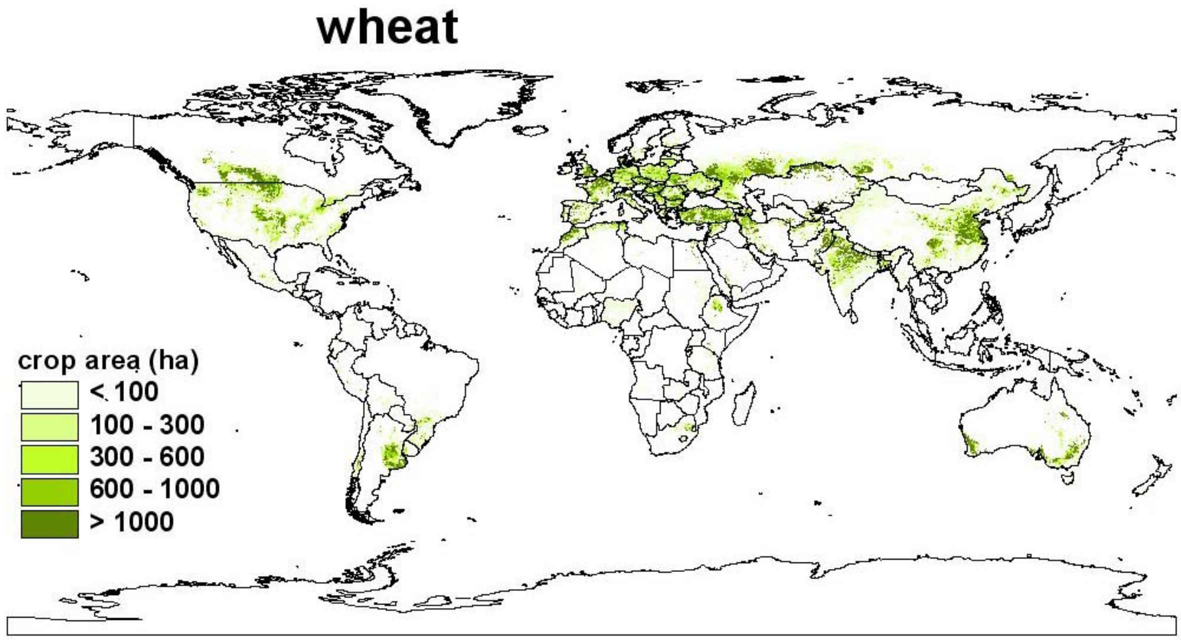
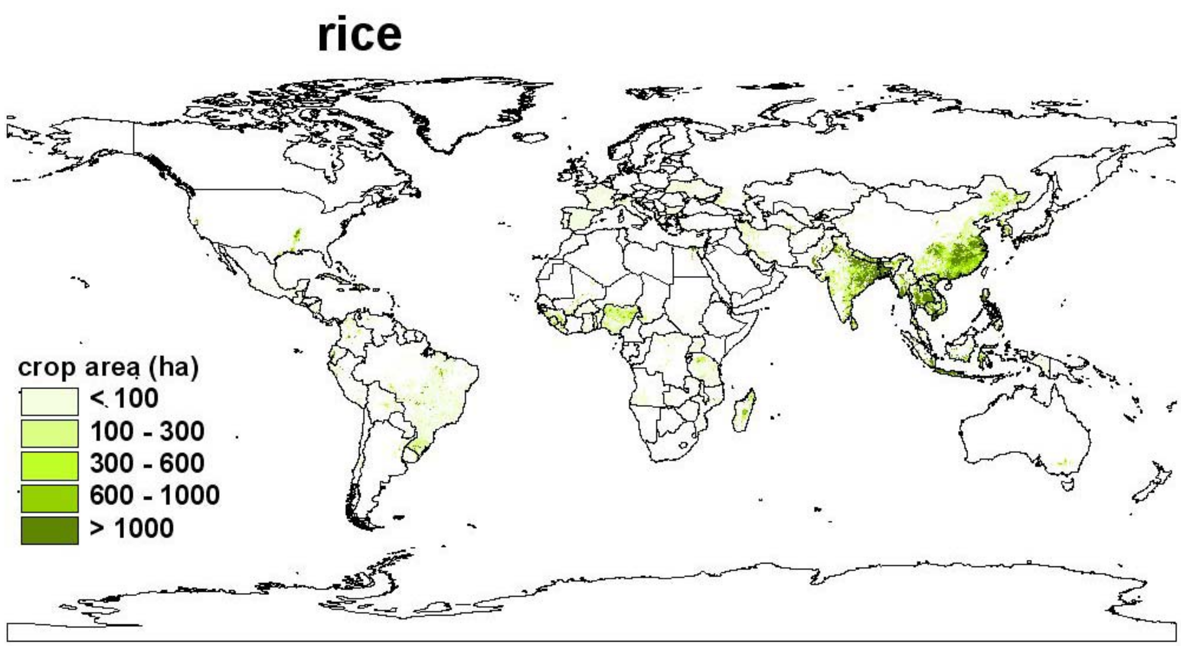
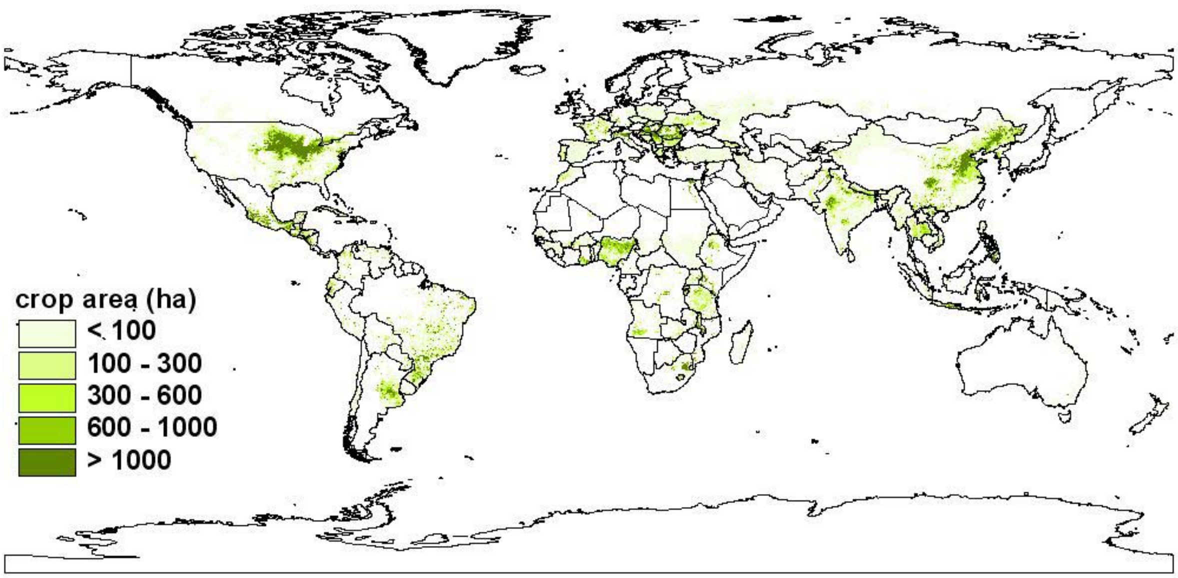


Figure 2 Estimated crop distribution maps of the world





maize



sorghum