Spatial Allocation of Agricultural Production

Using a Cross-Entropy Approach

Liangzhi You* and Stanley Wood

International Food Policy Research Institute

2033 K Street, NW

Washington, DC 20006

USA

Correspondence: Dr. Liangzhi You (L.YOU@CGIAR.ORG)

Phone: 001-202-862-8168; Fax: 001-202-467-4439

ABSTRACT

While production statistics are reported on a geopolitical – often national - basis we often need to know, for example, the status of production or productivity within specific sub-regions, watersheds, or agro-ecological zones. Such re-aggregations are typically made using expert judgments or simple area-weighting rules. We describe a new, entropy-based approach to the plausible estimates of the spatial distribution of crop production. Using this approach tabular crop production statistics are blended judiciously with an array of other secondary data to assess the production of specific crops within individual 'pixels' – typically 1 to 25 square kilometers in size. The information utilized includes crop production statistics, farming system characterization, satellite-based interpretation of land cover, biophysical crop suitability assessments, and population density. An application is presented in which Brazilian state level production statistics are used to generate pixel level crop production data for eight crops. To validate the spatial allocation we aggregated the pixel estimates to obtain synthetic estimates of municipio level production in Brazil, and compared those estimates with actual municipio statistics. The approach produced extremely promising results. We then examined the robustness of these results compared to short-cut approaches to spatializing crop production statistics and showed that, while computationally intensive, the cross-entropy method does provide more reliable spatial allocations.

JEL classification: C60; Q15; Q24

Key Words: Entropy, cross entropy, remote sensing, spatial allocation, production, crop suitability

1. Introduction and Rationale

Internationally comparable series of annual crop production data are available at a national scale from FAO and USDA. While very rich in their commodity coverage, these data give no clue as to the geographic distribution of production within country boundaries. Several (sub-)regional collection efforts have been made by centres of the CGIAR (e.g., Carter *et al*, 1992; CIAT, 1996; CIP, 1999; Ladha *et al*, 2000; ILRI, 2001; IFPRI, 2001), by FAO (Gommes, 1996), and by the Famine Early Warning System (FEWS) in parts of Africa (http://www.fewsnet.org/). With the exception of the on-going mandate of FEWS to compile sub-national agricultural production and market data in many parts of sub-Saharan Africa, all of these were limited, one time efforts. Such collections of sub-national data are much more limited both in country and time coverage. The enormous gaps in geographic, time period, and crop coverage are unlikely ever to be filled. But even when sub-national data are available, they are often still inadequate in terms of providing sufficiently detailed insights into the location of production. Obtaining sub-national agricultural production data for, say, Lampung province in Indonesia, the state of Rondonia in Brazil, or the Valle de Cauca department in Colombia, would still reveal nothing about spatial variability of production within those areas of many thousand square kilometres, and yet to compile all such data globally, or even regionally, represents a formidable data discovery and harmonization challenge.[1] To combat this situation, the spatial allocation approach described in this

[1] Another even more challenging approach is to compile regional estimates of the spatial variation of crop production from national agricultural surveys and censuses. While such surveys typically allow greater spatial resolution in the determination of crop production, the sampling frameworks employed still limit the spatial scale at which results can be generated within acceptable levels of statistical confidence. There are

paper attempts to generate plausible allocations of crop production to the scale of individual pixels (notionally of arbitrary scale but in this application of some 100 km^2), through judicious interpretation of all accessible evidence. If this can be done with some confidence, we remove one of the major analytical weaknesses of regional and global agricultural studies – the inability to objectively re-aggregate production statistics into any other geography than national (or even sub-national) administrative boundaries. This has been a thorn in the flesh of the many macro studies that set out to analyze production and productivity by agro-ecological zones or watersheds, e.g., the agricultural research priorities study of Davis, Oram, and Ryan (1987), the CGIAR's Regional AEZ strategies of the 1990s (TAC 1992), the global food perspective studies from FAO (Alexandratos 1996; Bruinsma 2003), IFPRI (Rosegrant et al, 2001), and agroecosystem assessments (Wood, Sebastian and Scherr, 2000).

While it is technically feasible to detect the locations of some types of crops or crop groups (paddy rice and orchards being prime examples) using high-resolution satellite imagery such as LANDSAT and SPOT, together with proper ground-truthing, national land cover mapping conventionally focuses on the delineation of "natural" ecosystems that are often easier to detect, and of more direct interest to the forestry, wildlife, or environmental agencies who usually commission this type of work. The global 1km land cover database (IGBP 1998) does contain some crop-specific agricultural land cover interpretations in its regional and pre-classified background data (the *Seasonal Land Cover Regions*), but they are few and inconsistently applied. This global dataset does, however, provide a relatively detailed picture of where

also many complexities in conversion of local measurement units into standard units of area and quantity, and such surveys are often made only once per decade.

(undifferentiated) croplands may be found, and this is the very set of spatial boundaries within which crop-specific analyses can take place.

[Figure 1 The task of spatial crop allocation]

Figure 1 shows, diagrammatically, the challenge faced by the spatial allocation approach. The bold closed-curve shows the boundary of the geographical area within which we wish to identify specific pixels where specific crops are grown. The geographical area in our case is the (geopolitical) statistical reporting unit (SRU) for which we have been able to obtain production statistics. These may be national, first administrative level, e.g. state, or second administrative level, e.g., counties. The SRU is divided into pixels whose actual sizes depend upon the spatial resolution of available spatial data such as land cover and crop suitability surfaces. Knowing the reported harvested areas and production of crops in this SRU, the task of the spatial allocation model is to simultaneously allocate these crop areas and productions into those pixels within the geopolitical boundary where individual crops are most likely to be found. Some pixels may be allocated no crops, some pixels might be allocated some share of the SRU total for a single crop, and some pixels may be allocated multiple crops. The approach allows any single pixel to be occupied by multiple crops simultaneously.

The paper is organized as follows. The next section describes the types of information we use in the spatial allocation process. Section 3 introduces the allocation methodology – the cross-entropy method. In this section we first introduce the entropy concept, then describe the spatial allocation model in detail. Section 4 applies our model to data compiled for Brazil, a very large and agroecologically diverse country. We describe the application of the model and evaluate the accuracy of the allocation results.

In addition, we also compare the current model with other simplified crop allocation methods. Section 5 discusses the results and describes ongoing efforts to further develop the spatial allocation model.

2. Information Used to Assess the Spatial Distribution of Crop Production

The goal of the allocation is to spatially disaggregate SRU tabular statistics and assign them to specific "pixels" within a gridded map of the SRU. The information used to guide the spatial allocation comes in various forms.

1. *Crop production statistics.* The data include the harvested areas, production, and average yield for each crop being included in the allocation exercise. This tabular data is derived from international or national sources (e.g., FAOSTAT for national SRUs, national statistical yearbooks for first administrative level SRUs, and agricultural surveys for second level SRUs)
2. *Production system structure.* Agricultural production is diverse in terms of farming technology and the farm size. Normally, commercial farmers use more and higher quality production inputs such as high-yielding varieties, irrigation, fertilizers, pesticides, credit and market information, while subsistence farmers often rely on traditional cultivars and less replenishment of nutrients. The intent of partitioning of crop production among the various major production system types is to provide the allocation model with some guidance as to the manner in which different crops are produced. High input systems will likely have higher yields and are likely found in more favorable areas. This set of information could be obtained (with much effort) from a mix of sources such as small-scale studies,

farming system studies, country reports, agriculture survey data and even expert opinion.

3. *Cropping intensity.* Cropping intensity is defined as the number of cropping within a year for a certain crop. Most production statistics report the crop areas in terms of area harvested. From a spatial allocation perspective and consistency with satellite image of land cover, we need to convert harvested to physical areas. For example an irrigated rice field may produce two crops per year. Thus the 100,000 hectares of harvested rice reported in the SRU statistics is obtained from just 50,000 hectares of land. Similarly, maize and beans may be grown in combination in a single season. Thus 50,000 hectares of maize-bean cultivation will produce 50,000 hectares of both maize and beans during a given year. Thus for each crop, for each major type of production system identified, we must assess the relevant cropping intensity.
4. *Cropland extent:* Satellite-based land cover imagery is a key input in guiding the spatial allocation of crops. Based upon the processed image classifications, we reclassify the imagery into cropland extent and non-cropland. This remote sensing imagery provides the most detailed spatial data of the agricultural production. By default we will only allocate crop production within the extent of cropland as depicted by satellite data.
5. *Crop Suitability.* To a significant extent, the patterns and intensities of crop production are determined by the biophysical and soil conditions such as local landscape, variation in radiation, temperature, humidity, and rainfall, the quality of soil, and the occurrence of frosts, floods and droughts. There are many ways to

assess and represent the biophysical suitability of crop production, from general suitability classes through to estimates of potential yield and suitable areas under the given set of conditions at any location. Initially, FAO developed crop-specific hand-drawn map of crop suitability classes using spatial data on major climate regimes and length of growing period (FAO, 1981). More recently these approaches have been extended to the generation of digital suitability surfaces in the form of potential crop yields and suitable areas (Fischer et al, 2000). These new assessments of crop suitability also take account of soil related and slope factors.

6. *Existing crop distribution maps.* Any existing digitized or mapped data of the spatial distribution of specific crop based on direct field observation is a very valuable information source. In the current approach, the priors are very important and the existing crop distribution maps, even if it is only partial in its geographic coverage, improve our prior knowledge of where crops might actually be grown. For example, there are dot maps of production areas with the size of dots representing the size the production. The challenge is to correctly interpret the data and convert dots representing certain hectares (e.g. 100, 500, 2000) of crop production to actual spatially referenced cropped area.

3. Cross Entropy Approach

All the above information can be brought to bear on the spatial allocation of agricultural production one way or the other. But we need an approach that can utilize all such information, but that recognizes that the information may be limited, partially correct, and sometimes conflicting. Golan, Judge and Miller (1996) proposed various estimation techniques based on the principles of entropy. The advantage of this approach is described by Zellner (1988) as satisfying the "information conservation principle", namely that the estimation procedure should neither ignore any input information nor inject any false information (Robinson et al, 2000). In this sense, the entropy approach is an efficient information processing procedure for the spatial allocation task. The spatial allocation model uses the cross entropy (CE) approach that allows for the inclusion of prior knowledge about crop distribution. Using this methodology to allocate area within any particular geopolitical domain, it is straightforward to apply constraints that allocated areas are non-negative and that they sum up to the total reported area of each crop. The approach is also flexible in supporting the inclusion of additional equality or non–equality constraints that are considered to reflect the conditions under which the crop allocation must be performed.

3.1 Information Entropy

The cross entropy formulation is based upon the entropy concept in information theory originated by Shannon (1948). For a given probability distribution $\{p_1, p_2, \ldots, p_k\}$, Shannon's information entropy (amount of information) is defined as

$$H(p_1, p_2, \ldots, p_k) = -\sum_{i=1}^{k} p_i \ln p_i \tag{1}$$

where $\ln 0=0$ by convention, which means zero probability yields zero information.

Jaynes (1957) proposed a principle of maximum entropy to identify an unknown distribution of probability from given moment constraints. Kullback (1959), Good (1963) introduced the notion of cross-entropy, *CE*, which is a measure of the discrepancy between the two probability distributions, say p_i and q_i.

(2)
$$
\begin{aligned}
CE(p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_k) &= \sum_{i=1}^{k} p_i \ln(p_i / q_i) = \sum_{i=1}^{k} p_i \ln p_i - \sum_{i=1}^{k} p_i \ln q_i \\
&= -H(p_1, p_2, \ldots, p_k) - \sum_{i=1}^{k} p_i \ln q_i
\end{aligned}
$$

The cross entropy minimization approach provides a model formulation in which the discrepancies between **p** and its prior, **q,** are minimized subject to certain constraints.

3.2 Spatial Allocation Model

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, $HarvestedArea_j$ for each crop into an equivalent physically cropped area, $CropArea_{j.}$, using cropping intensity.

(3a)
$$CropArea_j = HarvestArea_j / CroppingIntensity_j$$

Another fundamental feature of crop production is production system distinguished by different levels of farming technologies. Farmers adopt quite different farming technologies to produce the same crop from one location to another, and these technologies greatly affect the crop performance. In the current model, we disaggregate each crop (e.g. rice) into three distinctive “crops” by the level of inputs and management, namely, irrigated, high-input rainfed, low-input rainfed (e.g. irrigated rice, high-input rainfed rice and low-input rainfed rice). Irrigated production is that equipped with irrigation equipment, and normally using high-yield modern varieties. By high-input rainfed, we mean rainfed production that is based on high-yielding varieties with modern inputs such as mechanics, nutrients, chemical pest and disease control. This production is mainly market oriented. Low input rainfed production means that based on traditional cultivars and labor-intensive techniques with no applications of nutrients. This system is largely subsistence or smallholder based. Let s_{ijl} be the share of the cropped area of crop *j* at input leve *l* allocated to pixel *i,* and since $CropArea_j$ is the total physical area for crop *j*, the area allocated to pixel *i* for crop *j,* A_{ijl}, is

(3b) $$A_{ijl} = CropArea_j \times Share_{jl} \times s_{ijl}$$

where $Share_{jl}$ is the share of total physical area for crop *j* at input level *l* .

In general we have some prior knowledge on crop-specific area distributions. Let π_{ijl} be the prior area shares we know for pixel *i* and crop *j* at input level *l.* The prior can be based upon an examination of existing crop distribution maps or any other information deemed relevant. For example, one could estimate a prior on crop distribution based upon biophysical, soil, social-economic attributions. The minimum cross entropy model is to choose a set of area shares s_{ijl} , such that

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

subject to:

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

(6) $$\sum_j \sum_l CropArea_j \times Share_{jl} \times s_{ijl} \le Avail_i \quad \forall i$$

(7) $$CropArea_j \times Share_{jl} \times s_{ijl} \le Suitable_{ijl} \quad \forall i \forall j \forall l$$

(8) $$1 \ge s_{ijl} \ge 0 \quad \forall i, j, l$$

where:

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

j: *j = 1, 2, 3, …,* crop identifier within the allocation unit, and

l: l = *irrigated, rainfed-high input, rainfed-low input,* management and input level fro crops.

$Avail_i$: 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 from FAO/IIASA suitability surfaces as introduced in the previous section.

The objective function of the spatial allocation model is the cross entropy of area shares and their prior. Equation (5) is adding-up constraints for crop-specific areas. Equation (6) 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 (7) is the constraint that the allocated crop area cannot exceed what are suitable for the particular crop. The last equation, Equation (8) is basically the natural constraint of s_{ijl} as shares of total crop areas. As we can see, the essence of this classic CE approach is to use any and all sources of information to best guess where crops might actually grow. In this sense, the spatial allocation model finds a solution that is consistent with the mean value of the aggregated data. The criterion for choosing the solution (out of many possible solutions because the problem is under-determine) is to minimize the entropy-based divergence from the prior.

Obviously, an informative prior is quite important for the success of the model. If some coarse crop distribution maps such as dots maps exist, we will definitely use them as the prior. However, we don’t have such a luxury in most situations or we only have partial (both in terms of crops and geographic coverage) data. In such cases, we resort to some simplified methods of allocating crop production used in the past. The most common method uses other information layers such as total land area, cropland and crop suitable areas. We choose crop suitability surfaces by FAO/IIASA because only this dataset provides the crop-specific information. The potential yields for crop *j* at input level *l* and pixel *i, $Suitability_{ijl}$,* is normalized to produce the prior:

$$\text{(9)} \qquad \pi_{ijl} = \frac{Suitability_{ijl}}{\sum_i Suitability_{ijl}} \quad \forall j \forall i \forall l$$

However, the crop suitability algorithm is purely based on the biophysical conditions. This may not reflect the reality in the field. For example, some remote areas may be quite suitable for certain agricultural crops but have not cultivated yet. We overlay crop suitability surface with the population density map, and change the suitability to zero (not suitable) for those pixels where population density is either extremely low (or zero) or extremely high (urban areas). This would produce a more realistic prior. In addition, we may simply take the normalized rural population density as the prior for low input rainfed portion of the crop if we identify that crop to be a subsistence one.

4. Model Application

We apply the above model to Brazil. Brazil has a total land area over 8.5 million square kilometers, in which only 6% is cropland. It is rich in natural resources and biodiversity, heterogeneous in agroecological conditions with quite different farming systems within its boundary. The first level administrative region is called state, the third-level municipio. Though there are only 27 states, there are over 4490 municipios in Brazil, averaging over 160 municipios per state. The spatial resolution in the current application is 5 by 5 Arc minutes, and the grid cell (pixel) with this resolution is about

9*km* by 9*km* (around 8,500 hectares) at the equator, and the cell size varies depending upon the latitude of that pixel. Brazil is covered by over 100,000 pixels of that size. Figure 2 shows the cropland, population density and the suitability map of maize for Brazil in 1994. Agricultural land is expressed as the percentage of each pixel occupied by cropland, as shown in Figure 2(a). This agricultural extent is estimated from the 1-km resolution global land cover database developed by the EROS DATA Center of the U.S. Geological Survey (Wood, Sebastian and Scherr 2000; Ramankutty and Foley 1998). Based upon the agricultural extent as shown in Figure 2(a), we calculate the agricultural area in each pixel by taking account of the change of pixel area with latitude. Figure 2(b) is the population density map. We set the population density limits of 5 $persons/km^2$ and 500 $persons/km^2$ for possible crop growing areas. In other words, the land beyond this range would be either city or forestry with little agriculture. As pointed out in Section 2, FAO/IIASA's newly-developed crop suitability surfaces are rich sources of information on both potential yields and suitable areas for each commodity under different management/input assumptions. FAO/IIASA suitability surfaces are defined for five production system types for each crop: rainfed - high input, rainfed – intermediate input, rainfed – low input, irrigated – high input and irrigated – intermediate input. Corresponding to our model specification, we omit the two intermediate input classes and represent production by just three input classes, namely, high-input rainfed, low-input rainfed and high-input irrigated (referred to henceforth as irrigated). We defined suitable areas within each pixel as the sum of the following four suitability classes in the original FAO/IIASA suitability database: very suitable, suitable, moderate suitable and marginal suitable. Accordingly the yield is calculated as the area-weighted average of the yields in the above four suitable classes. As an example, Figure 2(c) shows the suitable areas of maize and Figure 2(d) the potential yield distribution of maize. These maps provide the spatial allocation model the critical information on total agricultural land, yield suitability, and suitable areas.

[Figure 2 Agriculture land, market accessibility and suitability maps of maize]

The following eight crops are included in the spatial allocation model for Brazil: rice, wheat, maize, cassava, potato, green bean and soybean. Collectively, these eight crops account for nearly one quarter of the value of Brazilian agricultural output in 2000, and nearly half of all crop output (Alston et al, 2000). The base year of the spatial allocation is 1994 in which the satellite image of land cover was taken. To avoid using an atypical year, we take the average of 1993-95 statistics as the data for 1994. The allocation units are the 27 states in Brazil. We start with the tabular harvested area by state for these eight crops as show in Table 1. To be consistent with the suitability data introduced in the above, we need to pre-process the harvest area into three categories for each crop: irrigated, high-input rainfed, and low-input rainfed. This is done from the information on Brazilian farming system as shown in Table 2. Table 2 shows the percentages for irrigated and high-input rainfed areas for all the eight original crops in the 27 states of Brazil. Obviously, the percentage of the low-input rainfed area is the residual of 100% minus the sum of the above two percentages. In addition, some crops, in particular irrigated ones such as rice, are multiple cropped in many regions. The physical crop area to be allocated is calculated by dividing the harvested area in each of the above three categories for each crop by its corresponding cropping intensity[2]. Treating the irrigated, high-input rainfed and low-input rainfed for each crop as individual “crops”(in the modeling sense), the total crop numbers for the current Brazil case is 24.

[Table 1 Harvested area by states in Brazil(1993-95)]

[Table 2 Farming systems in Brazil (1993-95)]

The spatial allocation model described in the above section is performed for every state in Brazil to simultaneously allocate all the 24 “crop” areas into the pixels across the entire state. GAMS (GAMS, 2002) is used to solve the optimization problem[3]. The output is the crop areas per pixel in Brazil. Figure 3 shows the resulting spatial allocation of the eight crops on pixel level for Brazil. The intent of the spatial allocation model is not to try to match the real-world pixel by pixel, but rather to derive a substantially more informative spatial allocation of crop-specific area (production) than what is provided by production statistics shown in Table 1. We can see from Figure 3 that there is quite large variation of crop distribution among and within the states. Potato and sorghum are quite spread out while the productions of soybean and maize are more concentrated. These pixel-level results can be aggregated into any larger spatial scale proper to the analysis, for example watershed or agro-ecological zones.

[Figure 3 Spatial distribution of crop areas]

---

[2] The eight crops are all single cropped in Brazil except that there are double croppings for all low-input bean production in Brazil and for low-input rainfed potato in three states (Minas Gerais, Paraná, Tocantins).

[3] The size of optimization problem is huge due to large number of pixels within a state, high-performance solver is needed. In the current optimization, we use GAMS newly-developed PATHNLP solver.

5. Model Validation and Comparison

To assess how well the model does the allocation, we aggregate the pixel level production obtained from the spatial allocation model into the 4490 municipios in Brazil. Then we compare these municipal production estimates with the actual municipal production statistics for Brazil. Among the current eight crops, we are able to collect the municipal statistics for maize, rice, wheat, bean, cassava, and soybean. Figure 3 shows the graphs for these six crops, in which the horizontal axis is the actual municipal statistics while the vertical axis is the estimates from the model. The spatial allocations for wheat, maize and bean match the municipal statistics very well, with R^2 values all greater than 0.50. The correlation coefficients for the above three crops are 0.65, 0.54. 0.53, respectively. In the graphics for these three crops, the points clearly cluster around the 45 degree line from the origin to the upright corner. For the other three crops, however, the data points are more dispersed over the whole graphical area. In particular, there are quite some points in the lower bottom for both cassava and rice, which means much less allocated areas than municipio statistics. The R^2 values for cassava, rice and soybean are 0.47, 0.43, 0.40, respectively.

[Figure 4 Correlation of municipal production statistics and predictions made from the spatial allocation model: Brazil 1993-95]

To investigate the reasons for the differences among crops, we need to consider the quality of our municipio statistics (though we are attempted to treat these data as the "truth"). As we all know, getting subnational production data for developing countries is quite a challenge, and the accuracy of such data set is often questionable because of weak local capacity. This applies to Brazil municipios statistics data we have here. In addition, we estimated the harvested areas for those municipios which we cannot collect data directly by interpolating from historical time series and country-level data. Because of the uncertainty on the quality of municipio statistics, we cannot conclude that spatial allocations for the other three crops are not accurate. On the other hand, the strong correlation between allocation and census data suggests that the statistics data for wheat maize and bean are in good quality and the spatial allocations are quite close to the reality. Even though we have doubts on the quality of the statistics data, we still regard these statistics as "the truth" for the sake of assessing the spatial allocation model. As illustrated in Figure 4, the performance of the model is reasonable good with all correlation coefficients around 0.5, depending on specific crops. There are several reasons for the different performances among crops. First, different crops present different difficulties for satellite to capture the crop images. For example most of bean and rice producers in Brazil are smallholders. These production scatters sparsely within large areas of forest. This presents quite a challenge both for satellite imaging and for the subsequent interpretation. Secondly, the suitability surfaces (yield and suitable area) may have different accuracies for different crops in terms of representing the reality on the site. The suitability information is the driving forces in the spatial allocation because it provides both the prior and the constraints in the model. Third, the extent of the mismatch between biophysical suitability and actual local production system may be different from crop to crop due to local cultivation history and tradition. In addition, relatively less data points (less municipios who produce wheat) may be of the factors contributing the much better correlation.

[Table 3 Comparison of the effectiveness of alternative spatial allocation methods]

Without a more elaborate model, we could apply some simplified methods to allocate crop production into more disaggregated units. The most common method is simply to apply area shares of other information layers such as land area, suitability surface, and cropland surface as weights to allocate the total production of SRU into the sub-regions within that SRU. Table 3 shows R^2 values of comparing the results of such simple methods with municipio statistics of Brazil. We use four different layers: physical land area, suitable area, cropland, and the joint of cropland and suitable area. Surprisingly, even simply distributing the crop production by the physical land areas by municipios could get reasonable results (with correlation coefficients around 0.2-0.4). The results from cropland surface lead to quite decent correlation coefficients, which implied cropland estimation from satellite image is relatively accurate. As expected, our cross-entropy model beats all the simplified approaches with highest correlation coefficients for all crops. On one hand, these results give us some general idea on what layers give "better information" to guide our spatial allocation process. On the other, we could not generalize these results because these results are obviously country-specific and depend on the particular agricultural farming system of that country. For example, the high coefficients from land area shares approach implied that Brazilian productions of these six crops are evenly distributed within each state, which may not be true everywhere else.

5. Final Remarks

We have proposed a spatial allocation model for crop production statistics based on a cross-entropy approach (CE). The approach utilizes information from various sources such as satellite imagery, biophysical crop suitability assessments, as well as population density, in order to generate plausible, disaggregated estimates of the distribution of crop production on a pixel basis. In the application of the spatial allocation model to Brazil, a comparison of actual municipio production statistics with synthetic municipio estimates - generated from pixel level disaggregating of state level statistics - yielded R^2 values between 0.4 and 0.65. 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. However, the CE results do appear to produce spatial patterns of crop production significantly better than any other commonly used shortcut methods.

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.

REFERENCES

Alexandratos, N. (ed.) (1995). *World Agriculture: Towards 2010*. Chichester, UK: John Wiley and Sons, and Rome: Food and Agriculture Organization of the United Nations.

Alston, J. M., P.G. Pardey, S. Wood, and L. You (2000). Strategic technology investments for LAC agriculture: a framework for evaluating the local and spillover effects of R&D. International Food Policy Research Institute, Washington, D.C.

Anderson, J. M., E. Hardy, J. Roach, and R. Witmer (1976). *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*. Washington, DC: US Government Printing Office (also available on-line at http://mac.usgs.gov/mac/isb/pubs/factsheets/fs18999.pdf).

Carter, S.E., L.O. Fresco, P.G. Jones, and J.N. Fairbain (1992). *An Atlas of Cassava in Africa: Historical, Agroecological, and Demographic Aspects of Crop Distribution*. Cali, Colombia: Centro Internacional de Agricultura Tropical.

CIAT (1996). *Digital Map Dataset for Latin America and the Caribbean*. Cali, Colombia: Centro Internacional de Agricultura Tropical.

CIESIN, IFPRI and WRI (2000). *Gridded Population of the World, Version 2 alpha*. Center for International Earth Science Information Network (CIESIN) Columbia University, International Food Policy Research Institute (IFPRI), and World Resources

Institute (WRI). Palisades, NY:CIESIN, Columbia University. Available on-line at: http://sedac.ciesin.org/plue/gpw.

CIP (1999), *Potato Distribution dataset*, personal communication with Robert Hijmans, Centro Internacional de la Papa, Peru

Davis, J.S., P.A. Oram and J.G. Ryan (1987). *Assessment of Agricultural Research Priorities: An International Perspective*. Australian Centre for International Agricultural Research and International Food Policy Research Institute.

Fischer, Guenther, M. Shah, H. van Velthuizen, F. Nachtergaele (2000). *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

GAMS, 2002. General Algebraic Modeling System. on-line at www.GAMS.com.

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

Gommes, A. (1996). *AGDAT*. Rome, FAO.

Good, I. J. (1963), Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Annals of Mathematical Statistics*, 34, 011-934

IFPRI(2001). Latin American Subnational Crop Production Database. International Food Policy Research Institute, USA

IGBP (1998). Global 1-km Land Cover Set DISCover, International Geosphere Biosphere Programme (IGBP) Data and Information System, IGBP-DIS (information available online at http://www.igbp.kva.se/lucc.html).

ILRI (2001). A Spatially-Referenced Crop and Livestock Production System Database for Eastern and Southern Africa. Project AS 9716. Nairobi. International Livestock Research Institute.

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

Jaynes, E. T. (1957b). Information Theory and Statistical Methods II, *Physics Review*, 108(1957): 171-190

Kullback, J. (1959). *Information Theory and Statistics*, New York: John Wiley & Sons

Ladha, J. K., K. S. Fischer, M. Hossain, P. R. Hobbs, and B. Hardy (eds) (2000), “Improving the Productivity and Sustainability of Rice-Wheat Systems of the Indo-Gangetic Plains: A Synthesis of NARS-IRRI Partnership Research”, *IRRI Discussion Paper No. 40*, Manila: International Rice Research Institute.

Ramankutty, N. and J.A. Foley (1998). Characterizing Patterns of Global Land Use: An Analysis of Global Croplands Data. *Global Biogeochemical Cycles* 12(4): 667-685.

Robinson, S., A. Cattaneo and M. El-Said (2000). Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods, *Trade and Macroeconomics Division Discussion Paper No.58,* International Food Policy Research Institute, Washington D.C.

Rosegrant, M. W., M. Pasiner, S. Meijer and J. Witcover (2001). *Global food projections to 2020: Emerging trends and alternative futures*. IFPRI 2020 Vision Report, International Food Policy Research Institute, Washington, DC, USA.

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

TAC/CGIAR (1992). *The Ecoregional Approach to Research in the CGIAR*. Report of TAC/Center Directors Working Group. TAC Secretariat, FAO, Rome.

Wood, S., 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 USA.

Zellner, A.(1988). Optimal Information Processing and Bayes Theorem, *American Statistician*, 42, 278-284

Figure 1 The task of spatial crop allocation

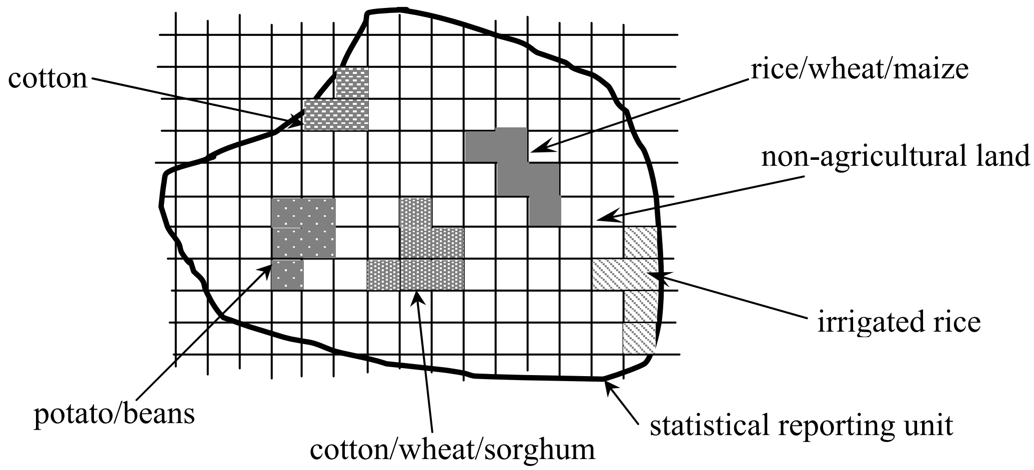

Figure 2 Agriculture land, population density and suitability maps of maize

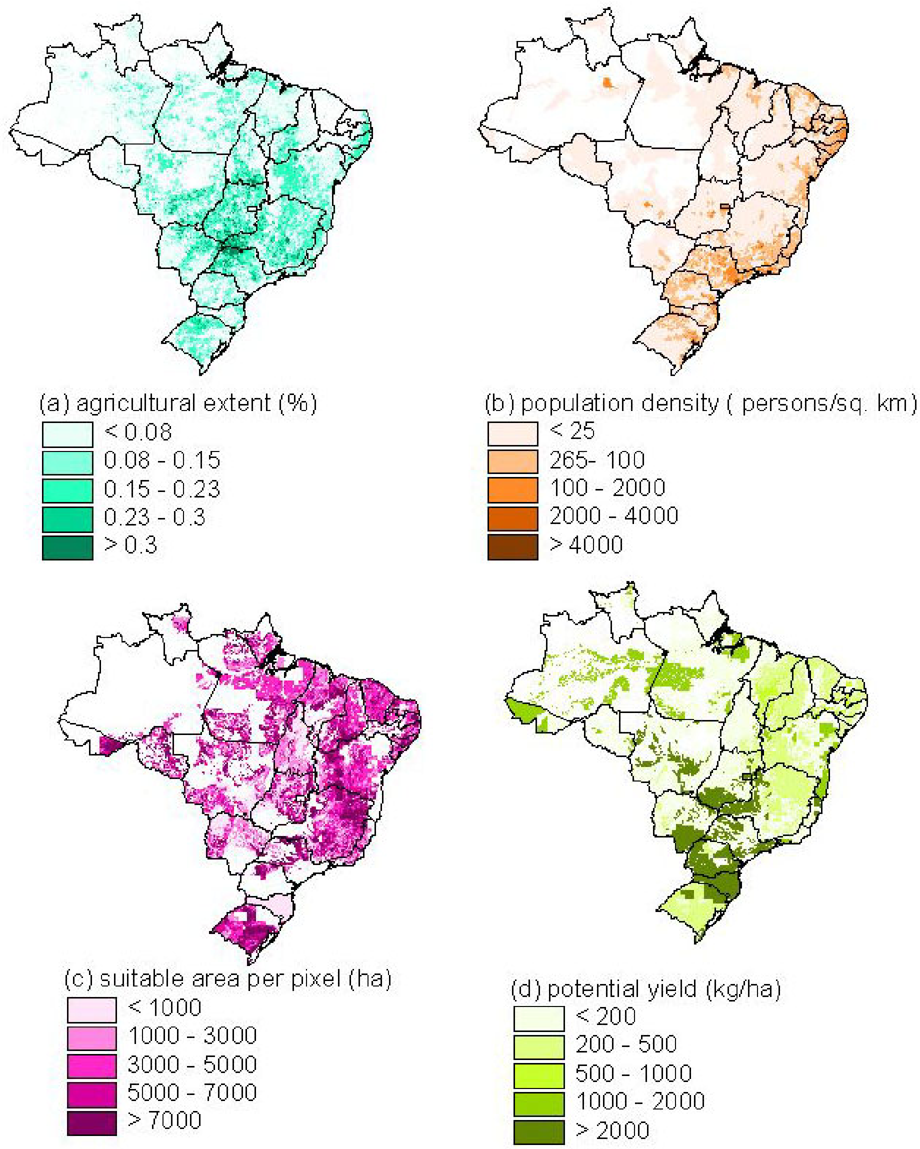

Table 1 Harvested areas by states in Brazil:1993-95

State	Wheat	Rice	Maize	Sorghum	Potato	Cassava	Bean	Soybean
				(Hectare)				
Brazil	**1,267,967**	**4,403,820**	**13,191,061**	**138,991**	**169,681**	**1,868,646**	**4,783,341**	**11,268,031**
Acre		34,051	36,402			22,270	15,256	
Alagoas		7,335	70,578			30,534	110,775	
Amapa		586				2,559		
Amazonas		3,223	4,423			34,930	2,672	
Bahia		57,089	428,296	17,929	1,433	249,724	583,680	428,119
Ceara		67,045	507,781	365	11	116,276	548,077	
District Federal	778	2,058	20,253	81	462	495	5,598	46,149
Espiritu Santo		26,576	100,869		646	20,621	67,139	
Goias	3,093	289,420	842,786	31,019	276	17,759	142,947	1,070,754
Maranhao		759,309	602,078			261,855	116,897	64,778
Mato Grosso		461,965	404,532	18,281		24,315	39,646	2,005,885
Mato Grosso do Su	48,360	99,552	410,283	1,042	11	27,570	35,778	1,069,634
Minas Gerais	4,110	375,871	1,487,266	9,417	30,773	77,313	531,982	580,839
Para		204,696	245,100			266,333	81,067	
Paraiba		7,180	173,552	24	957	41,987	192,298	
Parana	646,682	108,955	2,647,208	164	43,050	147,792	559,837	2,142,562
Pernambuco		5,284	232,759	962	249	85,630	261,661	
Piaui		269,344	401,136	12		94,623	288,078	7,286
Rio Grande do Nor		1,758	97,821	3,417		47,691	127,176	
Rio Grande do Sul	471,334	983,221	1,782,287	28,660	45,792	107,934	208,633	3,086,668
Rio de Janeiro		17,043	26,812		175	13,781	11,913	
Rondonia		143,690	193,290			38,175	147,854	4,861
Roraima		8,783	6,479			2,655	1,578	
Santa Catarina	58,227	149,866	1,040,708		18,947	53,200	354,897	213,873
Sao Paulo	35,382	146,696	1,300,673	27,618	26,842	31,996	279,462	524,341
Sergipe		6,466	56,828		57	40,754	59,541	
Tocantins		166,758	70,860			9,875	8,897	22,283

Source: IFPRI Subnational Data (2000), and EMBRAPA

Table 2 Farming system in Brazil 1993-95

State	Irrigated Area*								High-input Rainfed Area*							
	Wheat	Rice	Maize	Sorghum (%)	Potato	Cassava	Bean	Soybean	Wheat	Rice	Maize	Sorghum (%)	Potato	Cassava	Bean	Soybean
Acre											89			63		
Alagoas											26	100		16		
Amapa											86			63		
Amazonas											40			35		
Bahia		2	10	0	100				100		50	85		41	10	95
Ceara		16									40	89		40	0	
District Federal	100				70		23			100	98			54	15	100
Espiritu Santo					30						71			76	10	
Goias	60	2			70		29		40	20	98	100		87	14	99
Maranhao		1								99	39	100		20	0	97
Mato Grosso							20			70	97	100		74	0	99
Mato Grosso do Su		65							98	7	97			74	10	99
Minas Gerais	100	39			40		11			49	84	98	36	66	9	95
Para									80		86			63		
Paraiba											48	99		19		
Parana		19			80		5		90	24	71	91		53	19	97
Pernambuco					5						41	97		21	0	
Piaui		4								10	45	99		45	0	95
Rio Grande do Nor		76			50					24	76	99		36	10	
Rio Grande do Sul		79									59	98		43	0	
Rio de Janeiro		99			50				85		64	95		41	10	95
Rondonia									100		73	99		76		
Roraima		57									95	100		81		
Santa Catarina		92			60				78	2	58	90		42	20	90
Sao Paulo					70		24		96	100	91	99		65	15	97
Sergipe		70			100					30	56	98		12		
Tocantins		34									96	100		93		98

Source: compiled by authors from a variety of statistical sources and expert opinions

**Note: Balance of production shares from each state are included in "low input rainfed" system*

Figure 3 Spatial distributions of crop areas

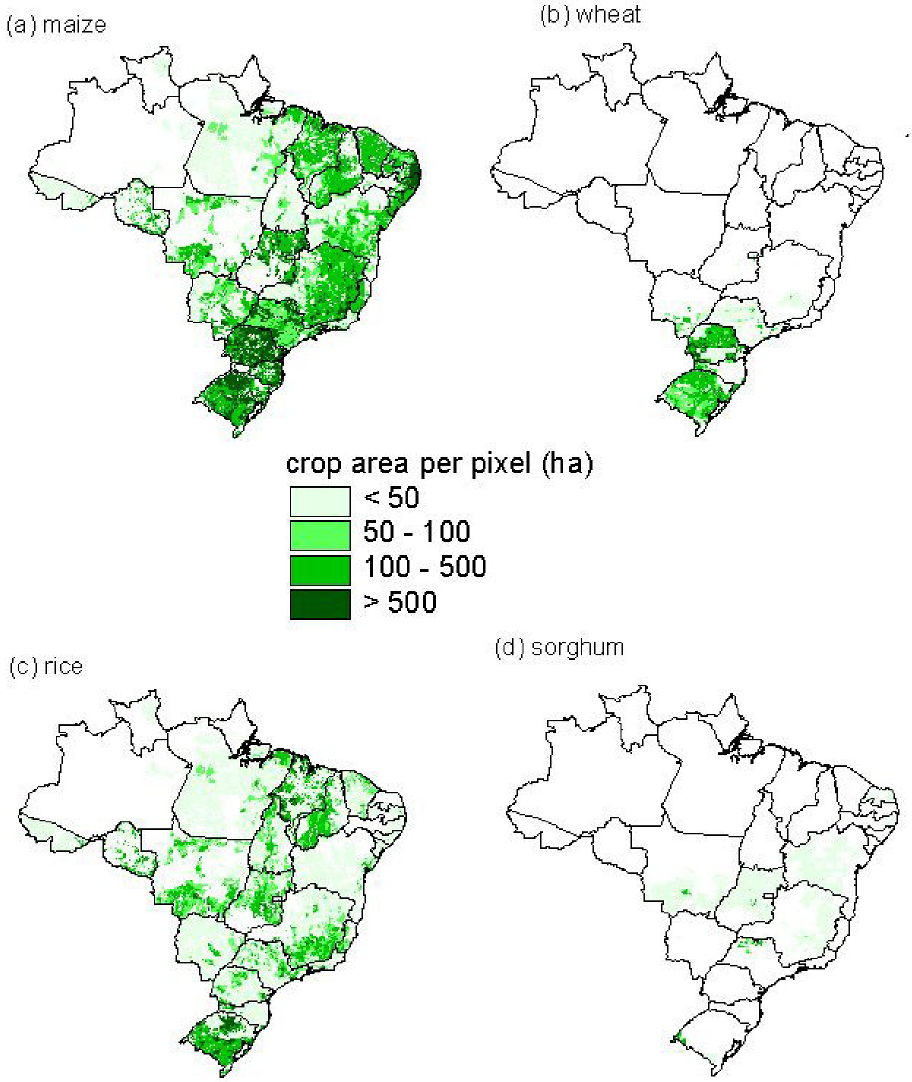

Figure 3 Spatial distribution of crop areas (continued)

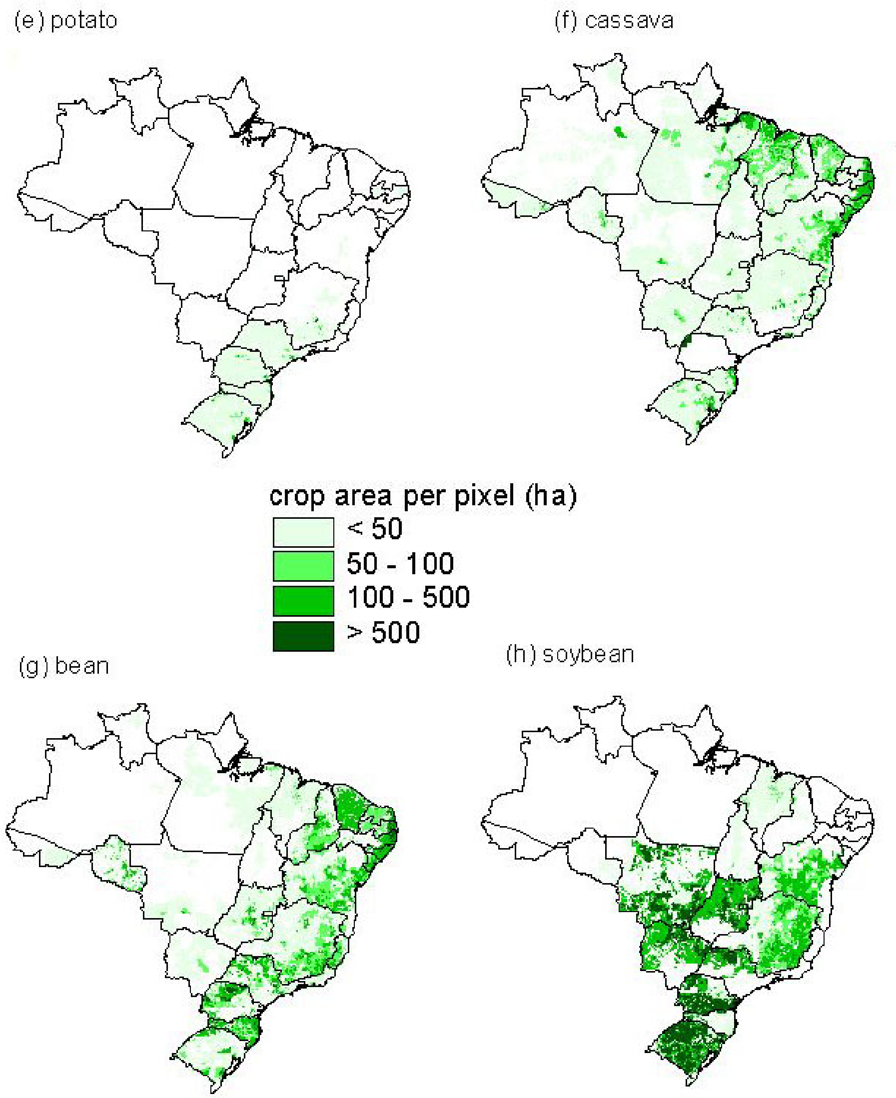

Figure 4 Correlation of municipal production statistics and predictions made from the spatial allocation model: Brazil 1993-95

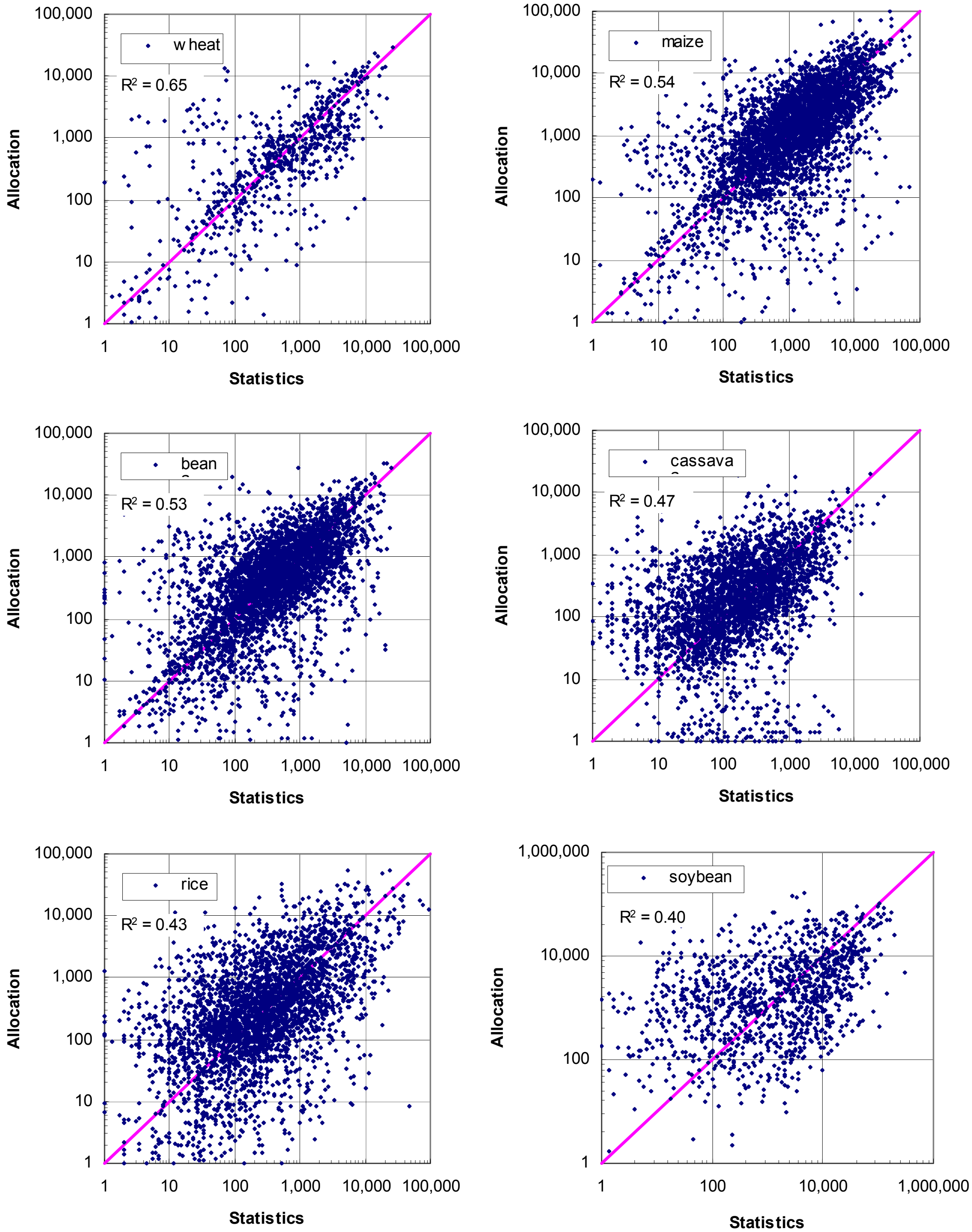

Table 3 Comparison of the effectiveness of alternative spatial allocation methods

Allocation Methods	Correlation Coefficients					
	wheat	rice	maize	cassava	bean	soybean
Land Area Shares	0.26	0.31	0.47	0.38	0.40	0.27
Suitable Area Shares						
Low Input	0.17	0.31	0.22	0.32	0.26	0.11
High Input	0.37	0.34	0.37	0.37	0.35	0.08
Irrigated	-0.04	0.32	0.01	0.45	0.13	-0.01
Mixed	0.15	0.38	0.17	0.39	0.28	0.04
Cropland Shares	0.38	0.31	0.44	0.38	0.25	0.37
Cross Entropy	**0.65**	**0.44**	**0.54**	**0.47**	**0.53**	**0.40**