ARTICLE IN PRESS

Agricultural Systems xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Agricultural Systems

journal homepage: www.elsevier.com/locate/agsy

Research Paper

Generating global crop distribution maps: From census to grid

Liangzhi You [a,b,*], Stanley Wood [c], Ulrike Wood-Sichra [b], Wenbin Wu [d]

[a] College of Economics and Management, Huazhong Agricultural University, Wuhan, Hubei 430070, China
[b] International Food Policy Research Institute, 2033 K Street, NW, Washington, DC 20006, USA
[c] Bill & Melinda Gates Foundation, 500 Fifth Avenue North, Seattle, WA 98109, USA
[d] Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China

ARTICLE INFO

Article history:
Received 10 February 2013
Received in revised form 11 January 2014
Accepted 11 January 2014
Available online xxxx

Keywords:
Global
Cross entropy
Spatial allocation
Agricultural production
Farming system
Crop suitability

ABSTRACT

We describe a new crop allocation model that adds further methodological and data enhancements to the available crop downscaling modeling. The model comprises the estimates of crop area, yield and production for 20 major crops under four rainfed and irrigated production systems across a global 5 arc minute grid. The new model builds on prior work by the authors (and published in this journal) in developing regional downscaled databases for Latin America and the Caribbean (LAC) and sub-Saharan Africa (SSA) and encompasses notions of comparative advantage and potential economic worth as factors influencing the geographic distribution of crop production. This is done through a downscaling approach that accounts for spatial variation in the biophysical conditions influencing the productivity of individual crops within the cropland extent, and that uses crop prices to weigh the gross revenue potential of alternate crops when considering how to prioritize the allocation of specific crops to individual grid cells. The proposed methodology also allows for the inclusion of partial, existing sources of evidence and feedback on local crop distribution patterns through the use of spatial allocation priors that are then subjected to an entropy-based optimization procedure that imposes a range of consistency and aggregation constraints. We compare the global datasets and summarize factors that give rise to systematic differences amongst them and how such differences might influence the fitness for purpose of each dataset. We conclude with some recommendations on priorities for further work in improving the reliability, utility and periodic repeatability of generating crop production distribution data.

1. Introduction

It is increasingly common for agricultural and environmental studies to rely on the use of gridded crop production data generated by the down-scaling of crop production statistics originally reported by more geographically-aggregated administrative units. The broad range of applications of such spatially down-scaled crop area, yield or production data includes climate change (Nelson et al., 2010; Lobell et al., 2008), food security (Hertel, 2011; Lobell et al., 2008), livestock production systems and systems evolution (Robinson et al. 2011, Herrero et al., 2010), technical change (Kostandini et al., 2009), ecosystem service valuation (Nelson et al., 2011), irrigation and rural road infrastructure (Dorosh et al., 2012; You et al., 2011), fertilizer input use (Liu et al., 2010).

The drive for improved spatial resolution of the location (area) and performance (yield) of crop production is fuelled by a number of mutually reinforcing factors. First and foremost, is growing awareness that a major obstacle to improving the effectiveness of policies and interventions aimed at improving rural well-being, agricultural growth, and natural resource sustainability is our inability to adequately account for the spatial heterogeneity of socio-economic, production, and environmental conditions (Nelson, 2002; Hertel, 2011). The more reliably we can assess the spatial distribution and covariance of such factors, the more cost-effective can be the formulation and targeting of appropriate policy and investment actions. Second, is the growing interest in understanding spatial patterns of agricultural production that might reveal untapped opportunities in, say, intensification and diversification, regional marketing, processing and trade or that might uncover significant levels of regional inequality and that, furthermore, might be helpful in shaping spatially-explicit strategic responses to such opportunities and challenges. Third, is simply the increasing ease and lower costs of exploring the spatial dimensions of agricultural development. Our capacity to acquire, manage, and share geo-referenced data has expanded significantly over the past twenty years, as have the range and utility of satellite and communications products and services – including the cropland

* Corresponding author at: International Food Policy Research Institute, 2033 K Street, NW, Washington, DC 20006, USA. Tel.: +1 202 862 8168; fax: +1 202 467 4439.

E-mail address: L.YOU@CGIAR.ORG (L. You).

http://dx.doi.org/10.1016/j.agsy.2014.01.002
0308-521X/

Please cite this article in press as: You, L., et al. Generating global crop distribution maps: From census to grid. Agr. Syst. (2014), http://dx.doi.org/10.1016/j.agsy.2014.01.002

and irrigated area land cover products utilized extensively in crop production down-scaling efforts such as those described here.

Within this broader context this paper describes a new global dataset that embodies significant methodological and data advances and that builds on regional approaches applied in Latin America and the Caribbean (You and Wood, 2006) and in sub-Saharan Africa (You et al., 2009). First we describe the global spatial production allocation model (SPAM) model in detail, the input data used to implement the model, and the generation of a new, global database of down-scaled crop production. Then we discuss the evaluation process we went through and the comparison of our result to the other two published, similar datasets. In the final section we reflect on the strengths and weaknesses of our approach and identify key areas in which progress is needed both in our own work as well as to better support the broader research and development community who strive for greater reliability, more frequent updates, and higher spatial resolution of down-scaled crop production data.

2. The spatial production allocation model

Following our previous work in America and the Caribbean (You and Wood, 2006) and sub-Saharan Africa (You et al., 2009), we further modified our entropy-based allocation model into a global SPAM model. The essential approach can be viewed as a triangulation or triage across a range of several of the most relevant sources of data known to represent factors that either are known to condition the likelihood of or to reflect the existence of certain crops.

2.1. Developing prior estimates of the spatial distribution of crop areas

An informed prior (π_{ijl}) is very important for the success of the model. We create the prior based upon the available input data. First for each pixel, we calculate the potential unit revenue of planting a certain crop as

$$\mathrm{Re}\,v_{ijl} = \mathrm{Price}_j \times Access_{ij} \times PotYield_{ijl} \tag{1}$$

where $Price_j$ and $Access_{ij}$ are the price and market accessibility indices for crop j at pixel i (accessibility only) for the statistical reporting unit (SRU). $PotYield_{ijl}$ is the potential yield for crop j at input level l and pixel i, which is described in the online supplementary material (S1). Market is important for both subsistence farmers and commercial ones. Even in poor countries where self-consumption is high, a large majority of households still purchase food products produced by others (p. 35, Losch et al., 2012). So many researchers have assumed that farmers are risk averse and profit maximizers (e.g. Hazell and Norton, 1986; Roundevell et al., 2003). We adopted this assumption in our model too. While the gross revenue could be reasonably estimated using Eq. (1), collecting cost data and so estimating profit on a global scale remains a daunting challenge if possible at all. Therefore, we rely on empirical evidence to further modify the revenue estimation. Firstly, subsistence farmers grow crops mainly for their own consumption and profit or even suitability may not be the determinant factor. So for subsistence part of a crop (l = *subsistence*), we simply use rural population density as a weight to pre-allocate the crop areas. $\bar{A}_{ijl}$ is the area pre-allocated to pixel i for crop j at level l:

$$\bar{A}_{ijl} = SubCropArea_{jkl} \times \frac{Pop_i}{\sum_{i\in k} Pop_i} \quad l = subsistence \quad \forall i \forall j \tag{2}$$

Secondly, revenue or even profit optimization alone could not explain the complex factors which determine farmers' production choices. For example, a certain regions produce more of a crop by historical or cultural reasons. Local demand and preference also strongly affect the production patterns. Let $Percent_{jl}$ be the area percentage of crop j at input level l of the total cropland in the SRU. We account for such non-economic factors by adding $Percent_{jl}$ into the revenue definition:

$$\mathrm{Re}\,v_{ijl} = Percent_{jl} \times \mathrm{Price}_j \times Access_{ij} \times PotYield_{ijl} \quad \forall l \neq subsistence \quad \forall i \forall j \tag{3}$$

Thirdly, Rev_{ijl} are modified to account for the evidence of existing crop distribution according to the likelihood of a certain crop's presence in the pixel. If the existing crop maps (See S1) or our evaluation process show crop j exists in pixel i, Rev_{ijl} is assigned an arbitrary big number (e.g. 5 times of the maximum revenue within the SRU); if crop j is shown very likely present at pixel i, Rev_{ijl} is set to the maximum value for crop j within the SRU. And so on and so forth. Setting higher values for Rev_{ijl} will force allocation to the pixels no matter what their calculated revenue values are.

Then we estimate the prior allocation of A_{ijl} using irrigated area, cropland and the above estimated revenue:

$$\bar{A}_{ijl} = IRRArea_i \times \frac{\mathrm{Re}\,v_{ijl}}{\sum_j \mathrm{Re}\,v_{ijl}} \quad \forall j \forall i \quad \forall l = irrigated \tag{4}$$

If $(Avail_i - IRRArea_i - \sum_j \bar{A}_{ij,subsistence}) \geqslant 0$, then

$$\bar{A}_{ijl} = (Avail_i - IRRArea_i - \bar{A}_{ij,subsistence}) \times \frac{\mathrm{Re}\,v_{ijl}}{\sum_j \sum_l \mathrm{Re}\,v_{ijl}} \quad \forall j \forall i \quad \forall l = ra\inf ed \tag{5}$$

Otherwise, $\bar{A}_{ijl} = 0, \forall i \forall j \quad \forall l = ra\inf ed$. What the above equations do is to breakdown the aggregated irrigated area and cropland (from satellite) into crop-specific areas, using the revenue as a weight. After this pre-allocation, we calculate the prior by normalizing the allocated areas over the whole allocation unit.

$$\pi_{ijl} = \frac{\bar{A}_{ijl}}{\sum_i \bar{A}_{ijl}} \quad \forall j \forall i \forall l \tag{6}$$

2.2. Spatial disaggregation crop production

Following our earlier work (You and Wood, 2006; You et al., 2009), we define our spatial crop allocation problem in a cross entropy framework. We first need to convert the allocated area into a probability value between 0 and 1. We accomplish that by using the area share allocated to pixel i and crop j at input level l within a statistical reporting unit (s_{ijl}). A SRU is normally a geopolitical unit such as country, state/province. $CropArea_{jl}$ is the total physical area of this SRU for crop j at input level l to be allocated. A_{ijl} is the area allocated to pixel i for crop j at input level l. Therefore:

$$s_{ijl} = \frac{A_{ijl}}{CropArea_{jl}} \tag{7}$$

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

$$\underset{\{s_{ijl}\}}{MIN} \quad CE(s_{ijl}, \pi_{ijl}) = \sum_i \sum_j \sum_l s_{ijl} \ln s_{ijl} - \sum_i \sum_j \sum_l s_{ijl} \ln \pi_{ijl} \tag{8}$$

subject to:

$$\sum_i s_{ijl} = 1 \quad \forall j \forall l \tag{9}$$

$$\sum_j \sum_l CropArea_{jl} \times s_{ijl} \leqslant Avail_i \quad \forall i \tag{10}$$

$$CropArea_{jl} \times s_{ijl} \leqslant SuitArea_{ijl} \quad \forall i \forall j \forall l \tag{11}$$