INTERNATIONAL FOOD
POLICY RESEARCH INSTITUTE
sustainable solutions for ending hunger and poverty
Supported by the CGIAR

IFPRI Discussion Paper 01327

February 2014

A Comparative Analysis of Global Cropping Systems Models and Maps

Weston Anderson

Liangzhi You

Stanley Wood

Ulrike Wood-Sichra

Wenbin Wu

Environment and Production Technology Division

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI), established in 1975, provides evidence-based policy solutions to sustainably end hunger and malnutrition and reduce poverty. The Institute conducts research, communicates results, optimizes partnerships, and builds capacity to ensure sustainable food production, promote healthy food systems, improve markets and trade, transform agriculture, build resilience, and strengthen institutions and governance. Gender is considered in all of the Institute's work. IFPRI collaborates with partners around the world, including development implementers, public institutions, the private sector, and farmers' organizations, to ensure that local, national, regional, and global food policies are based on evidence. IFPRI is a member of the CGIAR Consortium.

AUTHORS

Weston Anderson (W.Anderson@cgiar.org) is a senior research assistant in the Environment and Production Technology Division of the International Food Policy Research Institute (IFPRI), Washington, DC.

Liangzhi You (L.You@cgiar.org) is a senior research fellow in the Environment and Production Technology Division of IFPRI, Washington, DC.

Stanley Wood is a senior program officer in the Data and Diagnostics Agricultural Policy & Global Development Program of the Bill & Melinda Gates Foundation, Seattle, WA.

Ulrike Wood-Sichra is a research analyst in the Environment and Production Technology Division of IFPRI, Washington, DC.

Wenbin Wu is an employee of the Institute of Agricultural Resources and Regional Planning of the Chinese Academy of Agricultural Sciences, Beijing.

Notices

¹ IFPRI Discussion Papers contain preliminary material and research results and are circulated in order to stimulate discussion and critical comment. They have not been subject to a formal external review via IFPRI's Publications Review Committee. Any opinions expressed are those of the author(s) and do not necessarily reflect the policies or opinions of IFPRI.

² The boundaries and names shown and the designations used on the map(s) herein do not imply official endorsement or acceptance by the International Food Policy Research Institute (IFPRI) or its partners and contributors.

Copyright 2014 International Food Policy Research Institute. All rights reserved. Sections of this material may be reproduced for personal and not-for-profit use without the express written permission of but with acknowledgment to IFPRI. To reproduce the material contained herein for profit or commercial use requires express written permission. To obtain permission, contact the Communications Division at ifpri-copyright@cgiar.org.

Contents

Abstract	v
Acknowledgments	vi
1. Introduction	1
2. Global Cropping System Models	2
3. Comparison of the Downscaling Methodologies	4
4. Input Data and Model Interdependencies	6
5. Quantitative Comparison of Cropping System Maps	8
6. Conclusions	18
Appendix A: Model Methodology	19
Appendix B: Gaussian Filter Sensitivity Analysis	23
Appendix C: Cropland Extent, Harvested Area, and Yield by Crop and Model	25
References	32

Tables

2.1 Summary of products offered	3
4.1 Input data layers of the four models	6
A.1 Priorities in distributing the condensed crop calendars to monthly growing area grids	19

Figures

5.1 GAEZ and Ramankutty cropland extent by latitude	9
5.2 Cropland extent differences after applying a Gaussian filter with a kernel density of three sigma	9
5.3 Wheat harvested area and yield by latitude	10
5.4 Comparison of wheat harvested area by model following a Gaussian filter of three sigma kernel density	11
5.5 Comparison of wheat yield by model following a Gaussian filter of three sigma kernel density.	12
5.6 Rice harvested area and yield by latitude	13
5.7 Comparison of rice harvested area by model following a Gaussian filter of three sigma kernel density	14
5.8 Comparison of rice yield by model following a Gaussian filter of three sigma kernel density.	15
5.9 Comparison of maize harvested area by model following a Gaussian filter of three sigma kernel density	16
5.10 Maize harvested area and yield by latitude	17
5.11 Comparison of maize yield by model following a Gaussian filter of three sigma kernel density	17
B.1 Differences in cropland extent with Gaussian filters having kernel densities of 0 (pixel-level comparison), 1 (4-pixel radius), 2 (8-pixel radius), 3 (12-pixel radius), and 4 (16-pixel radius)	23
C.1 Cropland extent of (A) GAEZ and (B) Ramankutty et al. (2008)	25
C.2 Pixel-wise cropland extent differences	25
C.3 Wheat harvested area for (A) M3, (B) GAEZ, (C) MIRCA, and (D) SPAM	26
C.4 Pixel-wise comparison of the wheat harvested area by model	26
C.5 Wheat yield for (A) M3, (B) GAEZ, and (C) SPAM	27
C.6 Pixel-wise comparison of the wheat yield by model	27
C.7 Wheat harvested area for (A) M3, (B) GAEZ, (C) MIRCA, and (D) SPAM	28
C.8 Pixel-wise comparison of the rice harvested area by model	28
C.9 Rice yield for (A) M3, (B) GAEZ, and (C) SPAM	29
C.10 Pixel-wise comparison of the rice yield by model	29
C.11 Maize harvested area for (A) M3, (B) GAEZ, (C) MIRCA, and (D) SPAM	30
C.12 Pixel-wise comparison of the maize harvested area by model	30
C.13 Maize yield for (A) M3, (B) GAEZ, and (C) SPAM	31
C.14 Pixel-wise comparison of the maize yield by model	31

ABSTRACT

Agricultural practices have dramatically altered the Earth's land cover, but the spatial extent and intensity of these practices is often difficult to catalogue. Cropland accounts for nearly 15 million square kilometers of Earth's land cover—amounting to 12 percent of the planet's ice-free land surface—yet information on the distribution and performance of specific crops is often available only through national or subnational statistics. Although remote-sensing products offer spatially disaggregated information, those currently available on a global scale are ill suited for many applications due to the limited separation of crop types within the area classified as cropland. Recently, however, the field has seen multiple independent efforts to incorporate the detailed information available from statistical surveys with supplemental spatial information to produce a spatially explicit global dataset specific to individual crops for the year 2000. Although these datasets provide analysts and decision makers with improved information on global cropping systems, the final global cropping maps differ from one another substantially. This study aims to explore and quantify systematic similarities and differences between four major global cropping systems products: the dataset of monthly irrigated and rainfed crop areas around the year 2000 (MIRCA2000), the spatial production allocation model (SPAM), the global agroecological zone (GAEZ) dataset, and the M3 dataset developed by Monfreda, Ramankutty, and Foley. The analysis explores not only the final cropping systems maps but also the interdependencies of each product, methodological differences, and modeling assumptions, which will provide users with information vital for discerning between datasets in selecting a product appropriate for each intended application.

Keywords: global cropland, yield, harvested area, cropping systems, mapping, downscaling

ACKNOWLEDGMENTS

The authors would like to thank Dr. Van Velthuis and Dr. Fischer for their helpful correspondence regarding the GAEZ model; Dr. Portmann and Dr. Siebert for providing additional data from the MIRCA model; and Dr. Licker for providing the climate zone dataset used to calculate the yield gap. This paper would not have been possible without their generous contributions. The work presented here is partly sponsored by the Bill & Melinda Gates Foundation (HarvestChoice program), the US National Science Foundation (award ID: 1048967), and the National Basic Research Program of China (973 program: no. 2010CB951502). These supports are greatly appreciated.

1. INTRODUCTION

Cropland accounts for approximately 15 to 18 million square kilometers of Earth's land cover—amounting to 12 percent of the planet's ice-free land surface—yet information on the distribution and performance of specific crops is often available only through national or subnational statistics (Ramankutty et al. 2008; Foley et al. 2005). Cataloging the increasing extent and yield of cropland has implications for food security analyses, studies of land degradation, and resource management. Whereas subnational statistics provide information on raw quantities, they provide only limited information useful for spatially explicit applications. Detailed mapping of human-induced land-use change is vital to understanding water usage (Rosegrant, Cai, and Cline 2002), nutrient cycling (Bondeau et al. 2007; Liu et al. 2010), soil erosion (Yang et al. 2003), loss of biodiversity (Foley et al. 2011), and impact on regional climate (Ramankutty, Delire, and Snyder 2006; Voldoire et al. 2007).

Remote-sensing products offer spatially disaggregated information, but those currently available on a global scale are ill suited for many applications due to the limited separation of crop types within the area classified as cropland. Recently, however, several initiatives have begun to incorporate the detailed information available from statistical surveys with supplemental spatial information to produce a spatially explicit global dataset specific to individual crops for the year 2000. These studies have generated increasingly sophisticated results portraying the downscaling of crop production statistics across the global extent at a moderately high spatial resolution. Global studies have been reported by Leff, Ramankutty, and Foley (2004), Monfreda, Ramankutty, and Foley (2008), Portmann, Siebert, and Döll (2010), and most recently You et al. (2013) and Fischer et al. (2010). Although such datasets provide analysts and decisionmakers with improved information on global cropping systems, the final global cropping maps differ from one another substantially. This study aims to explore and quantify systematic similarities and differences between four major global cropping systems products and compare and contrast the general conceptual, methodological, and data underpinnings of the four products: the global dataset of monthly irrigated and rainfed crop areas around the year 2000 known as MIRCA2000 (Portmann, Siebert, and Döll 2010), the spatial production allocation model, or SPAM (You et al. 2013), the global agroecological zone (GAEZ) dataset (Fischer et al. 2013), and the M3 dataset developed by Monfreda, Ramankutty, and Foley (2008).

We begin with an overview of the methods and outputs of each dataset in Section 2 before comparing and contrasting the downscaling methodology each product uses in Section 3. Section 4 compares the input datasets each model uses and evaluates how interdependencies between models may propagate through the downscaling methodology to affect the final product. In Section 5, we perform a quantitative comparison across the four products, focusing on the world's three major crops: rice, wheat, and maize. We conclude with a summary and some recommendations for users of these products.

2. GLOBAL CROPPING SYSTEM MODELS

Research on cropping system models has been reported at the global scale by Leff, Ramankutty, and Foley (2004), You et al. (2013), Monfreda, Ramankutty, and Foley (2008), and Portmann, Siebert, and Döll (2010), while regional applications have been reported for Latin America and the Caribbean (You and Wood 2006) and Africa south of the Sahara (You, Wood, and Wood-Sichra 2009). This paper focuses on four global cropping system models: MIRCA, M3, GAEZ, and SPAM.

Monfreda, Ramankutty, and Foley's 2008 Cropping System Model (M3)

Of the four cropping system models considered, the M3 approach attempts spatial downscaling of the most complete coverage of crops (175, including tree and forage crops and managed grasslands) for both harvested area and yield. M3 is described in the companion paper to Ramankutty et al. (2008), which uses remote-sensing products to construct a new dataset for croplands and pasture circa 2000 at a 5-arc-minute resolution. The M3 dataset applies minimal modeling to distribute subnational statistics of yield and harvested area, opting for ease of interpretation and a limit to requisite assumptions over complexity (see Section 3 of Monfreda, Ramankutty, and Foley [2008] for further detail).

Dataset of Monthly Irrigated and Rainfed Crop Areas (MIRCA) around the Year 2000

MIRCA downscales 26 crops and two aggregate categories of “other annual” and “other perennial” crops, all of which are divided into rainfed and irrigated production areas. But, unique among the four approaches, MIRCA also performs a temporal downscaling so as to provide rainfed and irrigated area estimates disaggregated by month. MIRCA uses M3 downscaled crop data results as its starting point for allocating the total harvested area for each crop into rainfed and irrigated areas, and apportions the M3 crop area allocations into 402 spatial “calendar units” globally for which the MIRCA team has been able to compile unique sets of ancillary information on irrigation, crop-specific irrigated water use, crop calendars, and cropping intensities.

Spatial Production Allocation Model (SPAM)

SPAM covers the fewest crops, just 20, but downscales the area and yield for each crop into three different production systems: high-input irrigated, high-input rainfed, and low-input rainfed. The low-input rainfed category is itself further subdivided into low input and subsistence as described in You et al. (2013). SPAM relies on a separate collection of subnational statistical data from that of MIRCA, focusing on increased coverage in developing countries. The SPAM approach relies on using ancillary information—including crop prices, population density (CIESIN, IFPRI, and WRI 2000, and crop-specific biophysical suitability (Fischer et al. 2010)—to distribute subnational statistics based on a method known as cross entropy (Golan, Judge, and Miller 1996; Lencer and Miller 1998; Zhang and Fan 2001). See Section 2 of You et al. (2013) for detail on the cross-entropy approach.

Global Agroecological Zones Cropping System Model (GAEZ)

The GAEZ dataset (Fischer et al. 2010) downscales 23 crops, including forages and other cereals in either irrigated or rainfed production systems for both harvested area and yield. GAEZ develops a cropland extent independent from SPAM, MIRCA, and M3—which all use Ramankutty et al. (2008)—and relies on an extensive analysis of crop-specific agroclimatic and edaphic suitability. Using a methodology similar but not identical to SPAM, GAEZ incorporates ancillary information such as population density, biophysical suitability, and market access into a cross-entropy framework as a means of distributing subnational statistics.

Products Offered

Given their different approaches, the four datasets offer four different sets of products at the same 5-arc-minute resolution (about 9 kilometers at the equator) globally for the year 2000. As Table 2.1 shows, the final products of the four models are as follows:

- M3—350 global grids. One harvested area grid and one average yield grid for each of 175 crops.
- MIRCA—624 global grids. One harvested area grid for rainfed and one harvested area grid for irrigated production of 26 crops for each month.
- SPAM—240 global grids. One physical area grid, one harvest area grid, and one average yield grid for each of 20 crops for irrigated, high-input rainfed, and low-input rainfed (divided into low input and subsistence) production.
- GAEZ—138 global grids. One harvested area grid, one average yield, and one production value grid for each of 23 crops for irrigated and rainfed production systems.

Table 2.1 Summary of products offered

	M3	MIRCA	GAEZ	SPAM
<i>Crop system disaggregation</i>	175 crops (including forage crops and “grasses other”)	Rainfed grassland/pasture	23 crops (including forages and “other cereals”)	Managed grassland/pasture
		“Other annual crops”		“Other crops”
		26 irrigated crops + 26 rainfed crops (including forages, and “irrigated managed grassland/pasture”)		20 crops
	None	Irrigated, Rainfed	Irrigated, Rainfed	Irrigated, Rainfed (commercial), Rainfed (subsistence), Rainfed (subsistence low input)
<i>Production indicators</i>	Harvested area, Yield	Harvested area	Harvested area, Yield, Production value	Harvested area, Physical area, Yield
<i>Seasonality</i>	Annual	Monthly	Annual	Annual
<i>Global data products</i>	$175 \times 1 \times 2 \times 1 = 350$	$26 \times 2 \times 1 \times 12 = 624$	$23 \times 2 \times 3 \times 1 = 138$	$20 \times 4 \times 3 \times 1 = 240$

Source: Author’s calculations.

3. COMPARISON OF THE DOWNSCALING METHODOLOGIES

Cropland Extent Delineation

As a first step toward delineating crop-specific harvested area and yield, each cropping system model defined a spatially explicit layer of cropland extent, representing the proportion of cropland in each 5 arc-minute pixel globally. Because each subsequent step in the modeling process relies on the definition of cropland extent, the degree to which each pair of cropland extent products agree represents an upper bound of intermodel agreement on the spatial distribution of crop physical areas.

M3, MIRCA, and SPAM all rely on the same base dataset for cropland extent—Ramankutty et al. (2008), which is an extension of Leff, Ramankutty, and Foley (2004). Leff, Ramankutty, and Foley synthesize satellite-derived land cover data and agricultural census data worldwide to assess the distribution of major crops across a global 5-arc-minute grid in terms of the proportion of the total harvested area of each of the crops in each administrative unit. Following and improving on this work, Ramankutty et al. (2008) developed a new global land cover dataset for croplands and pasture circa 2000 (at the same 5-arc-minute resolution of the original dataset) by combining Boston University's MODIS-derived land cover data (Friedl et al. 2002) and the SPOT VEGETATION-based GLC2000 (Bartholome and Belward 2005). Ramankutty et al. (2008) apply a multiple linear regression model to relate the combined satellite-derived datasets to the agricultural statistics using a least-squares-error framework. The optimization is applied separately to six different regions of the world.

The cropland extent developed by Ramankutty et al. (2008) is used directly by M3 and with modifications by MIRCA and SPAM. By combining Ramankutty et al. and the Global Map of Irrigation Areas (GMIA), MIRCA produced a global dataset of monthly growing areas of 26 irrigated crops on the same 5-arc-minute grid. SPAM similarly reconciles the GMIA and the Ramankutty cropland extent by setting the cropland extent to be at least equal to the irrigated area in a preprocessing step. For more information on each of these methodologies, see Appendix A.

GAEZ uses GLC2000 data and GMIA, but it also considers a global land cover categorization (IFPRI 2002) that is based on a reinterpretation of the Global Land Cover Characteristics Database version 2.0 (EROS Data Center 2000), a layer of forestland from the *Forest Resources Assessment* of FAO (FAO 2001), the IUCN-WCMC protected areas inventory (WDPA 2009), and an estimate of land required for housing and infrastructure for the year 2000 derived from FAO-SDRN, based on LANDSCAN 2003 and calibrated to United Nations 2000 population figures (Fischer et al. 2008; Bhaduri et al. 2002; Dobson et al. 2003). GAEZ runs a cross-sectional regression on the land cover distributions to derive weights that are then applied in an iterative adjustment procedure to match estimated reference values such that the geographic and statistical data are consistent.

Suitability Constraints

GAEZ and SPAM further constrain potential crop distribution using biophysical and socioeconomic suitability prior to allocating the harvested area and yield of each crop. M3 and MIRCA do not consider suitability criteria. SPAM directly uses the suitable area product from GAEZ, meaning that despite using different cropland extent products to constrain the distribution of crops, the two models use identical constraints on biophysically suitable land. The GAEZ suitability product integrates an extensive set of edaphic and climatic factors into its biophysical suitability analysis to produce a suitability index by production system and crop. Further information on the suitability index analysis developed as part of the GAEZ model may be found in Fischer et al. (2013).

In addition to biophysical suitability criteria, both SPAM and GAEZ model the socioeconomic factors that often constrain or encourage crop production. As a means of differentiating between low, medium, and high input or management conditions, GAEZ divides the land into land-use types. Land-use types are derived using information on road infrastructure, livestock density, population density, and distance to market. For example, whereas low input relies on available human or livestock labor, high

input is market oriented. Similar to GAEZ, SPAM explicitly models different production systems, which include high-input irrigated, high-input rainfed, low-input rainfed, and subsistence (always low-input rainfed). SPAM also includes data on crop prices and market access to construct a realistic market scenario that includes not only biophysical barriers to producing crops but also social economic forces (see Appendix A for an explicit mathematical formulation).

Distribution of Harvested Area and Yield

Perhaps the largest methodological differences between M3, MIRCA, SPAM, and GAEZ lie in the approaches used to downscale statistical data reported at the administrative-unit level into grid-cell-specific values. M3 uses the most straightforward method, allocating each crop into each grid cell as the same proportion of grid cell cropland area as the crop occupies in the total harvested area of each statistical reporting unit. Crop yield in each grid cell is assigned as being the same as the yield reported for the statistical unit as a whole. This approach implicitly assumes that both environmental conditions and management/production systems are uniform across the cropland extents of each statistical reporting unit, or that there is insufficient information to characterize the spatial variations of crop production within a statistical unit. As a result, the distinct tolerances of individual crops to those spatial patterns are not incorporated in the downscaling procedure (for example, wetter, higher elevations within a province for which we hold statistical data may provide the production zones for cassava, banana, and coffee, while the lower elevations, which are drier and hotter, contain all the millet and sorghum areas). This approach does not, furthermore, acknowledge the very significant differences between the yield levels of irrigated and rainfed production systems, or of commercial and smallholder producers within these sometimes large and highly diverse statistical reporting units.

MIRCA primarily focuses on reconciling the differences among information derived from subnational crop production statistics, M3 crop distributions, and the Siebert et al. (2005) irrigated areas database. MIRCA deals only with harvested area and essentially uses the relative share of rainfed and irrigated cropland within each grid cell to break out M3 total crop areas into grid-cell-specific rainfed and irrigated areas. The model also includes use of numerous checks and adjustments to reconcile differences between the cropland area of Ramankutty et al. (2008) and the irrigated area estimates of Siebert et al. (2005) within each grid cell, given that total cropland area should at all times be greater than or equal to the irrigated cropland area. Similar to M3, MIRCA does not consider any form of suitability in its downscaling procedure.

The downscaling approaches of GAEZ and SPAM are predicated on the importance, in terms of subsequent utility of the downscaled estimates, of attempting to take explicit account of available evidence of the spatial variation of production conditions within the cropland extent and of the significantly different yield levels of different types of production systems even in the same production environment. Both GAEZ and SPAM use an approach that produces a result mathematically equivalent to that of a cross-entropy formulation, but GAEZ uses an iterative rebalancing procedure to adjust weighting factors until all constraints in the model are met, while SPAM uses a cross-entropy formulation. Although the two models incorporate similar information (see Table 4.1), the manner in which the information is used to constrain the model differs (see Appendix A for details on the mathematical formulation of each model). Additionally, GAEZ differs from SPAM in that it uses a *location factor* to incorporate spatially explicit information including geo-referenced household survey data. Although the prior in SPAM is used to capture spatially explicit information as well, the model does not include household survey data but instead leverages the field presence of the CGIAR network to incorporate an extensive dataset of expert elicitations.

4. INPUT DATA AND MODEL INTERDEPENDENCIES

The major determinants of the potential reliability of downscaling efforts are (1) the quality of the cropland extent dataset indicating the physical extent and area intensity of cropland (for example, share of cropland area in each 5-arc-minute grid cell), and (2) the resolution and reliability of the subnational crop statistics. Each model builds on a common set of available data as well as previous work in cropping systems modeling. Table 4.1 illustrates both the broad linkages and increasing sets of input data and assumptions that each of the M3, MIRCA, GAEZ, and SPAM datasets relies upon.

Table 4.1 Input data layers of the four models

Category	Dataset		M3	MIRCA	SPAM	GAEZ
National/Sub national statistics	FAOSTAT – National land use and crop production stats		x	x	x	x
	FAO AGRO-MAPS – Subnational crop statistics (SAGE and IFPRI collaboration)*		x	x	x	x
Cropland extent and cropping intensity	FAO AQUASTAT – National irrigation crop statistics			x	x	x
	GIMA – Global irrigated land			x	x	x
	GAEZ – Multicropping index		x	x	Data on crop production system	x
	GLC2000		x	M3 cropland extent modified*	M3 cropland extent modified*	x
	Boston University MODIS-derived land cover		x	M3 cropland extent modified*	M3 cropland extent modified*	
	GAEZ sustainability index				x	x
Ancillary data	Population density	FAO-SDRN population density (derived from LANDSCAN 2003)			x	x
		GRUMP population density				
	FAO ruminant livestock density					x
	Distance to market					x
	Crop prices				x	
	Expert elicitation and ancillary data collection			x	x	No documentation available

Source: Author's calculations.

Note: *MIRCA 2000 relies on M3 as does GAEZ “for selected crops in countries where more than 50% was covered by subnational statistics.”

National and Subnational Statistics

All four datasets draw on FAOSTAT national data to provide control totals for cropland area, the harvested area, and yields of specific crops, while also spending considerable efforts to collect subnational crop statistics to allow as detailed as possible disaggregation of national totals within subnational administrative boundaries. Since MIRCA relies on M3 to provide its input data on the spatial allocation of the total area and average yield (that is, not yet disaggregated between rainfed and irrigated production), it relies initially on the same sources of subnational crop statistics. SPAM relies on a separate collection of subnational statistical data sources, focusing on increased coverage in developing countries.

M3 reports a total of 22,106 statistical reporting units globally, of which 56 were national, 2,299 were first-level subnational disaggregation (for example, US state level), and 19,751 were second-level (for example, US county level) reporting units. SPAM reports 24,507 statistical units of which 251 were national, 2,758 were first level, and 21,498 were second level. SPAM focused its data collection efforts particularly in developing countries. For example, in Africa the M3 and SPAM datasets were developed using around 300 and 4,150 second-level statistical reporting units, respectively. The GAEZ model uses data from FAOSTAT to constrain the model at a national level and, similar to MIRCA, uses the M3 subnational statistics for select crops in countries that had subnational statistics covering more than 50 percent of the country.

Extent of Irrigation

Those models that distinguish between rainfed and irrigated cultivation—MIRCA, SPAM, and GAEZ—all use GMIA version 4.0 released in 2007 (Siebert et al. 2005) to identify the location and area intensity of irrigated production. However, the MIRCA and SPAM teams compiled information in the national and subnational shares of different production systems and cropping intensities independently. MIRCA and SPAM both draw on FAO's AQUASTAT and national databases for gaining greater insights into national and crop-specific irrigation extents and practices, but MIRCA relies on a richer collection of national data, including a more complete collection of national/subnational crop calendars and cropping intensities (in part because the goal of MIRCA is to produce monthly and not annual crop distribution maps). In contrast to MIRCA and SPAM, GAEZ relies on International Institute for Applied Systems Analysis AEZ data for information about cultivation intensity of irrigated crops.

Ancillary Data

SPAM and GAEZ incorporate datasets beyond those used by M3 and MIRCA as a means of differentiating between production levels within cropping systems. The SPAM approach requires additional sets of data because it attempts further disaggregation of its rainfed production statistics among commercial and subsistence categories, and bases its approach to distribution of individual crops within the cropland extent on agronomic, economic, and demographic principles and assumptions. These include crop area and production shares among irrigated production and large-scale/commercial and smallholder rainfed production, the spatial differences in the biophysical suitability of individual crops for irrigated and rainfed (commercial and subsistence) production, and estimates of the spatial patterns of population density as well as crop prices. GAEZ similarly divides the land into land-use types to reflect variable management and input conditions. Data used to differentiate among land-use types reflect the specific requirements of each and include road infrastructure, livestock density, population density, and distance to market. Table 4.1 reflects the overlapping and separate ancillary datasets used in SPAM and GAEZ.

In addition to available ancillary datasets, SPAM leverages the international network and field presence of CGIAR to undergo a systemic validation process. The feedback from this validation is used to inform future model simulations. This process is unique to the SPAM model.

5. QUANTITATIVE COMPARISON OF CROPPING SYSTEM MAPS

Methods

Each model studied produced spatially explicit cropping system maps on the same 5-arc-minute grid; however, comparing those grids directly (pixel-wise comparison) may produce artificially inflated disagreement between products. Using a pixel-specific approach fails to account for the spatial dimension of the data and implicitly assumes each pixel to be a result independent from any neighboring pixels.

Each product was therefore assessed using methods that incorporated the spatial dimension of the data in biophysically and mathematically meaningful ways. As a means of accounting for the biophysical evolution of crops and cropland by growing region while still allowing for methodological differences in crop distribution, the sum of crops or cropland for each product was compared by latitude. This analysis of each product provides a biophysically meaningful broad-brush-stroke supplement to the subsequent evaluation of the distribution of crops and cropland in both spatial dimensions. To compare the two-dimensional distribution of each product to one another, a Gaussian filter with a kernel density of three standard deviations was applied to the results of each product prior to a pixel-wise comparison. Preprocessing the data using a filter expands the analysis to incorporate neighboring pixels. The kernel density for the Gaussian filters—that is, the number of neighboring pixels to consider—was chosen following a sensitivity analysis using a kernel density of one, two, three, and four standard deviations. The results of the sensitivity analysis and full documentation of the implementation of the Gaussian filter are documented in Appendix B.

The pixel-wise, by-latitude, and Gaussian filter analyses were each applied to assess the cropland extent, the harvested area, and the yield for each product. SPAM, M3, and MIRCA all rely on the Ramankutty cropland, while GAEZ has developed its own cropland extent. The cropland extent delineates the domain to which each product's downscaling approaches are applied and therefore represents an upper limit to the agreement between products. Each model produced maps of harvested area for wheat, rice, and maize. But only M3, GAEZ, and SPAM produced maps of yield. The results of each comparison are described in following sections.

Cropland Extent

Although both Ramankutty et al. (2008) and GAEZ use the GLC2000 land cover dataset as one input to the definition of cropland extent, Ramankutty et al. blend GLC2000 with remote-sensing observations from the Boston University MODIS dataset while GAEZ supplements the GLC2000 data with independent information on the extent of protected areas, forests, and agricultural extent (Table 4.1) (Ramankutty et al. 2008; FAO 2001; WDP 2009; Friedl et al. 2002). The methodological differences in delineating cropland extent result in distributions of cropland that broadly resemble one another but that differ significantly over select regions. By latitude, the two cropland extents largely agree over the majority of potential cropland—north of 15° N—but the distributions contain significant discrepancies further to the south. These discrepancies are particularly pronounced in the ranges of 5° to 15° N, 10° to 20° S, and 35° to 40° S (Figure 5.1).

Figure 5.1 GAEZ and Ramankutty cropland extent by latitude

Source: Author's calculations.

The pixel-wise and Gaussian filter analyses provide greater detail on the source of the discrepancies identified in Figure 5.1. The GAEZ and Ramankutty cropland extent maps largely agree in Europe, southern Africa, East Africa, and through much of China. The products significantly disagree in the Great Plains of North America, West Africa, Southeast Australia, India, and Southeast South America (see Figure 5.2). These differences, which occur in the first step of the crop distribution process, propagate through each model to underpin the difference between products of those models using the Ramankutty et al. (2008) versus the GAEZ cropland, as discussed in the following sections.

Figure 5.2 Cropland extent differences after applying a Gaussian filter with a kernel density of three sigma

Source: Author's calculations.

Wheat Harvested Area and Yield

The differences in the harvested area of wheat between GAEZ and the products that use the Ramankutty cropland extent (Figure 5.3) largely mirror the differences in cropland extent depicted in Figure 5.2. The differences in the harvested area of wheat for Australia, the United States, and Russia are nearly identical to those same differences in cropland extent. The lack of disagreement between products in Africa and Brazil despite large discrepancies in cropland extent reflects how little wheat is grown in those areas, and is not an indication of model agreement (Figure 5.4).

Figure 5.3 Wheat harvested area and yield by latitude

Source: Author's calculations.

Differences between the M3 and MIRCA products are minor, which is to be expected given that MIRCA uses M3 output directly as the model's input (Table 4.1 and Figure 5.4, panel e). The difference between SPAM and M3 (and MIRCA) is much larger, in particular in Europe, North India, and coastal China. The large difference is a reflection of their different downscaling methodology and the different subnational data collections.

Figure 5.4 Comparison of wheat harvested area by model following a Gaussian filter of three sigma kernel density

Source: Author's calculations.

Note: Histograms in each panel display the normalized percent of pixels as a function of harvested area, y-axis limits (0, 50%), x-axis limits (-5000, 5000) hectares.

The estimated yields from M3, GAEZ, and SPAM (MIRCA does not produce estimates of yields) matched less well than did the harvested areas from each product. GAEZ predicted larger yields at higher latitudes, while M3 predicted a far larger wheat yield in the tropics than either SPAM or GAEZ (Figure 5.3). Judging from the spatial distribution, the anomalous mid-latitude wheat yield in M3 is predominantly in Africa, where M3 is alone in predicting significant yields over much of the continent (Figure 5.5). Similarly, in South America, M3 predicts more wheat across much of the continent than either of the other two models. GAEZ predicts wheat at much higher latitudes than SPAM or M3, stretching up into Canada and northern Europe. GAEZ also predicts concentrated yields in central/eastern China that neither of the other products predict.

Figure 5.5 Comparison of wheat yield by model following a Gaussian filter of three sigma kernel density

Source: Author's calculations.

Note: Histograms in each panel display the normalized percent of pixels as a function of yield, y-axis limits (0, 50%), x-axis limits (-20, 20) metric tons/hectare.

Rice Harvested Areas and Yields

The harvested areas of rice appear to be less influenced by discrepancies in the cropland extent products and instead differ as a function of downscaling method or input data. The spatial characteristics of the differences between products using Ramankutty versus GAEZ cropland extent do not mirror the differences in cropland extent, which reflects the fact that cropland extent of GAEZ and Ramankutty agree relatively well in the major rice-producing areas of the world when compared to those areas that produce wheat (Figures 5.2 and 5.4).

Evaluating the harvested areas of rice by latitude for each product reveals that MIRCA predicts significantly more harvested area for rice north of 30° N than do the other products (Figure 5.6). This discrepancy may stem from the fact that the MIRCA method of crop distribution does not consider biophysical limitations, or it may reflect differences in the input statistical crop yields at a subnational level given that the above-average estimation by MIRCA appears to be concentrated in eastern China (Figure 5.7).

With the exception of MIRCA's large harvested area in eastern China, the relation between GAEZ and each of the products that use the Ramankutty cropland extent (M3, MIRCA, and SPAM) is nearly identical (Figure 5.7), which may indicate that all three use similar subnational rice data in China. M3, MIRCA, and SPAM differ from GAEZ in their distribution of rice within India: GAEZ distributes more rice area to the southwest while other products distribute more rice to the northeast.

The rice yields match even less well than did the wheat yields, showing significant discrepancies between the products over all latitudes (Figure 5.6). M3 again predicts yields higher than do SPAM and GAEZ over most latitudes. Similar to the wheat yields, M3 consistently predicts greater yields in Africa than either of the other products as well as significantly higher yields in western South America and in north-central Asia (Figure 5.8). The differences between M3 and SPAM are most pronounced in Europe and Asia. SPAM predicts consistently higher yields across all of Europe while GAEZ predicts higher yields in India, China, and Southeast Asia.

Figure 5.6 Rice harvested area and yield by latitude

Source: Author's calculations.

Figure 5.7 Comparison of rice harvested area by model following a Gaussian filter of three sigma kernel density

Source: Author's calculations.

Note: Histograms in each panel display the normalized percent of pixels as a function of harvested area, y-axis limits (0, 50%), x-axis limits (-5000, 5000).

Figure 5.8 Comparison of rice yield by model following a Gaussian filter of three sigma kernel density

Source: Author's calculations.

Note: Histograms in each panel display the normalized percent of pixels as a function of yield, y-axis limits (0, 35%), x-axis limits (-20, 20) tons/hectare.

Maize Harvested Area and Yield

As with the harvested area of wheat, the discrepancies in the harvested areas of maize relate closely to differences in the cropland extent products in many regions. This relation is particularly apparent in South America, Africa, and North America (Figure 5.9). This suggests that once again, the underlying cropland extent dataset has a large effect on the final distribution of crops.

Figure 5.9 Comparison of maize harvested area by model following a Gaussian filter of three sigma kernel density

Source: Author's calculations.

Note: Histograms in each panel display the normalized percent of pixels as a function of harvested area, y-axis limits (0, 35%), x-axis limits (-5000, 5000)

The maize yields show consistently different patterns but broadly agree on the distribution by latitude. M3 does not display consistently higher yields, but does so in the extra-tropical latitudes (Figure 5.10). Those differences largely stem from maize yields in Chile, South Africa, Canada, and northern Europe/Russia (Figure 5.11). M3 also predicts significantly higher yields than either SPAM or GAEZ in the eastern United States.

Figure 5.10 Maize harvested area and yield by latitude

Source: Author's calculations.

Figure 5.11 Comparison of maize yield by model following a Gaussian filter of three sigma kernel density

Source: Author's calculations.

Note: Histograms in each panel display the normalized percent of pixels as a function of yield, y-axis limits (0, 35%), x-axis limits (-20, 20) metric tons/hectare

6. CONCLUSIONS

This paper explores and quantifies the systematic similarities and differences between four major global cropping systems modeling frameworks and their products. While the models have some similarities (for example, use the same input data) and interdependence (for example, MIRCA builds upon M3 and both GAEZ and SPAM use crop suitability), the modeling methods and their final products vary considerably. We found that the differences between the Ramankutty cropland extent and the GAEZ cropland extent manifest themselves in the final products, as demonstrated by the consistent discrepancies between the GAEZ harvested area products and the Ramankutty harvested area products (those of M3, MIRCA, and SPAM). When considering each harvested area product, regardless of the specific crop, the M3 and MIRCA datasets produce the most similar harvested area maps, which is to be expected because MIRCA uses M3 as a starting point. Furthermore, these models take the least complex modeling approach to distributing the statistical cropping data. Finally, we conclude that the harvested area products matched one another more closely than did the yield products, which differed significantly even in basic spatial patterns.

There are many reasons why the differences among SPAM, M3, MIRCA, and GAEZ are generally large. As this paper demonstrates, the most important reason lies in the input data and the methodologies used in the four models. M3 relies heavily on cropland extent while MIRCA relies mainly on the irrigated area and water use statistics of FAO's AQUASTAT and GMIA. The SPAM and GAEZ approaches synthesize the various data sources (satellite-based land cover, ground-based data, and modeling results) in an attempt to best estimate the crop production distribution. Although each model has its own weaknesses and strengths, our past experiences (for example, You and Wood 2006) demonstrate that relying on only one input layer alone (either cropland, crop suitability, or irrigated area) may work under certain circumstances but may not always be sufficient. Conversely, more input data and increasingly complex modeling does not necessarily lead to better or more accurate results.

In this paper we evaluate the extent to which the four models agree, and explore the root causes of their discrepancies. As the true crop distribution is unknown, we could not judge which product is more accurate than the other. Rather, such comparison gives the reader some sense of discrepancy among the four products and provides information for users to make a knowledgeable choice.

Moving forward, room certainly exists for the four modeling teams to collaborate and develop some community of practice. This has started in a recent workshop convened by IFPRI, but more needs to be done. In the meantime, the four products provide users with a range of alternative approaches to modeling cropping systems. Potential users of these cropping system products need to understand the input data and modeling differences prior to choosing the model best fit for their purposes. Although each model approaches the same problem using many of the same underlying datasets, the methods employed and assumptions made in each product significantly affect the final cropping system map.

APPENDIX A: MODEL METHODOLOGY

M3 Distribution of Crop Harvested Areas and Yields

In each grid cell that had agricultural inventory data, the map of crop area was calculated as follows:

$$fcrop_i = fcropland_i \left(\frac{crop_k}{cropland_k} \right), \quad (A1)$$

where $fcrop_i$ is the harvested area of a specific crop in pixel i , $fcropland_i$ is the fraction of pixel i designated as cropland, $crop_k$ is the harvested area of a specific crop in statistical reporting unit k , and $cropland_k$ is the amount of cropland in statistical reporting unit k . Yield was distributed uniformly across each grid cell as equivalent to the yield reported in the statistical reporting unit as a whole.

MIRCA Distribution of Harvested Area

MIRCA primarily reconciles the differences between the Siebert et al. (2005) dataset of areas equipped for irrigation (AEI), the cropland extent (CE) of Ramankutty et al. (2008), and the harvested area (HA) maps of the M3 dataset (Monfreda, Ramankutty, and Foley 2008) to provide a monthly cropping map for irrigated and rainfed crops. Table A.1 outlines the priorities used to reconcile inconsistencies between the datasets. MIRCA first produces a condensed crop calendar of harvested area for each subcrop c and Statistical Reporting Unit (SRU) k ($HAccc_{c,k}$). A subcrop is used to represent multicropping systems or different subgroups of a crop that grow at different points in the year. For a complete description of how the condensed cropping calendar is produced, see Portmann et al. (2008).

Table A.1 Priorities in distributing the condensed crop calendars to monthly growing area grids

Priority	Dataset	Goal
1	Area equipped for irrigation (Siebert et al. 2005)	In each month and grid cell, the sum of crop-specific areas is lower than or equal to the area equipped for irrigation.
2	Cropland extent (Ramankutty et al. 2008)	In each grid cell and month, the sum of crop-specific irrigated and rainfed areas is lower than or equal to the cropland extent.
3	Harvested crop area (Monfreda, Ramankutty, and Foley 2008)	In each grid cell and for each crop class, the annual sum of the irrigated and rainfed harvested crop area is equal to the total harvested area of the specific crop.

Source: Author's compilation.

Irrigated Crops

The MIRCA process is a production-system-specific cell-wise approach to disaggregating harvested area by month. Area equipped for irrigation and cropland extent both include fallow land in their definition, so these classes need not be used completely so long as the annual harvested area is disaggregated and designated as either rainfed or irrigated. Area equipped for irrigation is prioritized over cropland extent and harvested area in the following process:

1. Calculate the irrigated harvested area (IHA) for subcrop c in cell i of month m :

$$IHA_{c,i,m} = \frac{(HA_{c,i,m} \times fAEI_{i,m})}{\sum_{subcrop\ types_{i,m}}}, \quad (A2)$$

where $fAEI_{i,m}$ is the fraction of pixel i equipped for irrigation in month m .

2. Assign irrigated harvested areas to the monthly minimum of area equipped for irrigation and harvested area for subcrop c in cell i of month m :

$$\text{if } AEI > 0 \text{ and } CE > 0 \text{ and } HA_{c,i,m} > 0, \quad IHA_{c,i,m} = \min(AEI_{c,i,m}, HA_{c,i,m}). \quad (A3)$$

3. If there still exists $HAccc_{ck}$, distribute irrigated growing areas to those cells that have cropland extent greater than zero and that are equipped for irrigation even if no $HA_{c,i,m}$ exists:

$$\text{if } AEI > 0 \text{ and } CE > 0, \quad IHA_{c,i,m,MIRCA} = \text{remaining } HAccc_{ck}. \quad (A4)$$

4. If there still exists $HAccc_{ck}$, distribute it to areas within cells that are equipped for irrigation, even if the cropland extent (and therefore $HA_{c,i,m}$) is zero:

$$\text{if } AEI > 0 \text{ and } CE = 0, \quad IGA_{c,i,m,MIRCA} = \text{remaining } HAccc_{ck}. \quad (A5)$$

Rainfed Crops

Following the distribution of irrigated crop areas, rainfed crops were distributed. Rainfed annual crops were treated differently than rainfed permanent crops. Annual crops were allowed to grow on areas equipped for irrigation so long as they were available, while permanent crops were not.

5. Calculate the rainfed harvested area for crop c in pixel i of month m ($RHA_{c,i,m}$) by distributing rainfed crops to areas in which available cropland extent exceeds available area equipped for irrigation:

$$\text{if } CE > AEI, \quad RHA_{c,i,m} = \min(HA_{c,i,m}, CE_i). \quad (A6)$$

6. If there still exists $HAccc_{ck}$, expand suitable areas in cells with more cropland extent than area equipped for irrigation to 95 percent of the cell, leaving room to account for infrastructure:

$$\text{if } CE > AEI, \quad RHA_{c,i,m,MIRCA} = \min(HA_{c,i,m}, (0.95 * Area_i)). \quad (A7)$$

7. If there still exists $HAccc_{ck}$, expand suitable areas in cells with either cropland extent or area equipped for irrigation to 95 percent of the cell, leaving room to account for infrastructure:

$$\text{if } CE > 0 \text{ or } AEI > 0, \quad RHA_{c,i,m,MIRCA} = \min(HA_{c,i,m}, (0.95 * Area_i)). \quad (A8)$$

The total harvested area of all rainfed and irrigated crops is therefore the sum of the IHA calculated in steps A2 through A5 and the RHA calculated in steps A6 through A8.

SPAM Harvested Area and Yield Distribution

The SPAM model distributes available crop statistics using a cross-entropy approach that incorporates ancillary data on crop price, market access, biophysical suitability, and expert elicitation.

Harvested Area Calculation

SPAM distributes statistical information from allocation unit (for example, a country or a province) by using the cropping intensity of crop j in production system l to convert the reported harvested areas ($HA_{j,l}$) to physical areas ($CropArea_{jl}$) as follows:

$$CropArea_{jl} = \frac{HA_{jl}}{CroppingIntensity_{jl}} \quad \forall j, l. \quad (A9)$$

SPAM next defines the area allocated to pixel i for crop j in production system l (A_{ijl}) using the share of the total physical area for crop j in production system l ($Share_{jl}$) and the physical area ($CropArea_{jl}$) as follows:

$$A_{ijl} = CropArea_{jl} \times Share_{jl} \times s_{ijl} \quad \forall i, j, l. \quad (A10)$$

The minimum cross-entropy approach employed by the SPAM model calculates the area shares for crop j of pixel i in production system l as follows:

$$\min_{\{s_{ijl}\}} [CE(s_{ijl}, \pi_{ijl}) = \sum_i \sum_j \sum_l s_{ijl} \ln(s_{ijl}) - \sum_i \sum_j \sum_l s_{ijl} \ln(\pi_{ijl})], \quad (A11)$$

subject to the following constraints:

$$\sum_{i \in k} s_{ijl} = 1 \quad \forall j, i, k \quad (A12)$$

$$\sum_j \sum_l CropArea_{jl} \times Share_{jl} \times s_{ijl} \leq CroplandExtent_i \quad \forall i \quad (A13)$$

$$CropArea_{jl} \times Share_{jl} \times s_{ijl} \leq CropSuitableArea_{ijl} \quad \forall i, j, l \quad (A14)$$

$$\sum_{i \in k} \sum_l CropArea_{jl} \times Share_{jl} \times s_{ijl} = SubCropArea_{jk} \quad \forall k, j \in J \quad (A15)$$

$$\sum_{l \in L} CropArea_{jl} \times Share_{jl} \times s_{ijl} \leq AEI_i \quad \forall i \quad (A16)$$

$$1 \geq s_{ijl} \geq 0 \quad \forall i, j, l, \quad (A17)$$

where l may be irrigated, rainfed high-input, rainfed subsistence, or rainfed low input. $CroplandExtent_i$ is the total extent of cropland for pixel i , and $CropSuitableArea_{ijl}$ is the area suitable for crop j at input level l in pixel i . $SubCropArea_{jk}$ is the crop area statistics for crop j in subnational SRU k . AEI_i is the area equipped for irrigation in pixel i . J is a set of commodities for which subnational production statistics exist, and L is a set of commodities within pixel i that are irrigated. π_{ijl} represents the prior estimate of area shares for crop j at input level l in pixel i .

The prior is developed using expert elicitation where available and elsewhere is calculated based on potential unit revenue, Rev_{ijl} :

$$Rev_{ijl} = Share_{jl} \times Price_j \times Access_{ij} \times SuitableYield_{ijl}, \quad (A18)$$

where $Price_j$ is the price of crop j , and $Access_{ij}$ is a measure of the physical accessibility of the market for crop j from pixel i . $SuitableYield_{ijl}$ is the agroclimatically suitable yield for crop j at input level l in pixel i . Then the prior allocation of crop area is estimated using irrigated area and cropland as follows:

$$PriorArea_{ijl} = AEI_i \times \frac{Rev_{ijl}}{\sum_j Rev_{ijl}} \quad \forall j, i, \quad \forall l = irrigated. \quad (A19)$$

$$\begin{aligned} PriorArea_{ijl} = & (CroplandExtent_i - AEI_i - PriorArea_{ij, subsistence}) \\ & \times \frac{Rev_{ijl}}{\sum_l \sum_j Rev_{ijl}} \quad \forall j, i, \quad \forall l = rainfed. \end{aligned} \quad (A20)$$

In the case of subsistence farming, the revenue measure is replaced by a measure of population density. The subsistence part of the subnational crop area is then pre-allocated using rural population density as a weight.

$$PriorArea_{ij,subsistence} = SubCropArea_{jk} \times Percent_{jl} \times \frac{Pop_i}{\sum_{i \in k} Pop_i} \quad \forall j, i, l = \text{subsistence}. \quad (A21)$$

After this pre-allocation, the prior is calculated by normalizing the allocated areas over the whole allocation unit:

$$\pi_{ijl} = \frac{PriorArea_{ijl}}{\sum_i PriorArea_{ijl}} \quad \forall i, j, l. \quad (A22)$$

Yield Calculation

The calculation of yield is based on the statistical yield information for crop j within production system l for each SRU k . First, the average potential yield, $\bar{Y}_{jl}$, is calculated as follows:

$$\bar{Y}_{jlk} = \frac{\sum_i Suitability_{ijl} \times A_{ijl}}{\sum_i A_{ijl}} \quad \forall k, \quad (A23)$$

$$Y_{ijl} = \frac{Suitability_{ijl} \times CropYield_{jlk}}{\bar{Y}_{jlk}}, \quad (A24)$$

where $CropYield_{jlk}$ is the statistical yield reported for crop j in production system l within SRU k . Then the production of crop j in production system l and pixel i , $Prod_{ijl}$, could be calculated as

$$Prod_{ijl} = (A_{ijl} \times CroppingIntensity_{jl}) \times Y_{ijl}. \quad (A25)$$

GAEZ Distribution of Harvested Area and Yield

Similar to SPAM, the GAEZ model distributes available statistical data using a cross-entropy approach that incorporates ancillary data and expert opinion. The ancillary data used in the model include distance to market, population density, ruminant livestock density, and expert opinion. Although no detailed information about the mathematical formulation of the objective function or prior was available for GAEZ, details on the constraints used in the model may be found in the GAEZ version 3 documentation (Fischer et al. 2013).

APPENDIX B: GAUSSIAN FILTER SENSITIVITY ANALYSIS

The two-dimensional Gaussian filter for pixel i may be expressed as

$$g(x, y)_i = \left(\frac{1}{2\pi\sigma^2} \right) \left(e^{-\frac{x^2+y^2}{2\sigma^2}} \right), \quad (\text{B1})$$

where x is the distance from the horizontal axis, y is the distance from the vertical axis, and σ is the standard deviation of the Gaussian distribution, used to control the kernel density as illustrated below.

Figure B.1 Differences in cropland extent with Gaussian filters having kernel densities of 0 (pixel-level comparison), 1 (4-pixel radius), 2 (8-pixel radius), 3 (12-pixel radius), and 4 (16-pixel radius)

Source: Author's calculations.

Figure B.2 Differences in cropland extent with Gaussian filters having kernel densities of 0 (pixel-level comparison), 1 (4-pixel radius), 2 (8-pixel radius), 3 (12-pixel radius), and 4 (16-pixel radius)

Source: Author's calculations.

APPENDIX C: CROPLAND EXTENT, HARVESTED AREA, AND YIELD BY CROP AND MODEL

Figure C.1 Cropland extent of (A) GAEZ and (B) Ramankutty et al. (2008)

Source: Author's calculations.

Figure C.2 Pixel-wise cropland extent differences

Source: Author's calculations.

Figure C.3 Wheat harvested area for (A) M3, (B) GAEZ, (C) MIRCA, and (D) SPAM

Source: Author's calculations.

Figure C.4 Pixel-wise comparison of the wheat harvested area by model

Source: Author's calculations.

Figure C.5 Wheat yield for (A) M3, (B) GAEZ, and (C) SPAM

Source: Author's calculations.

Figure C.6 Pixel-wise comparison of the wheat yield by model

Source: Author's calculations.

Figure C.7 Wheat harvested area for (A) M3, (B) GAEZ, (C) MIRCA, and (D) SPAM

Source: Author's calculations.

Figure C.8 Pixel-wise comparison of the rice harvested area by model

Source: Author's calculations.

Figure C.9 Rice yield for (A) M3, (B) GAEZ, and (C) SPAM

Source: Author's calculations.

Figure C.10 Pixel-wise comparison of the rice yield by model

Source: Author's calculations.

Figure C.11 Maize harvested area for (A) M3, (B) GAEZ, (C) MIRCA, and (D) SPAM

Source: Author's calculations.

Figure C.12 Pixel-wise comparison of the maize harvested area by model

Source: Author's calculations.

Figure C.13 Maize yield for (A) M3, (B) GAEZ, and (C) SPAM

Source: Author's calculations.

Figure C.14 Pixel-wise comparison of the maize yield by model

Source: Author's calculations.

REFERENCES

- Bartholome, E., and A. S. Belward. 2005. "GLC2000: A New Approach to Global Land Cover Mapping from Earth Observation Data." *International Journal of Remote Sensing* 26: 1959–1977.
- Bhaduri, B. L., E. A. Bright, P. R. Coleman, and J. E. Dobson. 2002. "LandScan: Locating People Is What Matters." *Geoinformatics* 5 (2): 34–37.
- Bondeau, A., P. C. Smith, S. Zaehle, S. Schaphoff, W. Lucht, W. Cramer, and B. Smith. 2007. "Modelling the Role of Agriculture for the 20th Century Global Terrestrial Carbon Balance." *Global Change Biology* 13 (3): 679–706. doi: 10.1111/j.1365-2486.2006.01305.x.
- CIESIN (Center for International Earth Science Information Network), IFPRI (International Food Policy Research Institute), and WRI (World Resources Institute). 2000. Gridded Population of the World (GPW) dataset, version 2. CIESIN, Columbia University, Palisades, New York. Accessed on August 1, 2013. <http://sedac.ciesin.columbia.edu/plue/gpw>.
- Dobson, J. E., E. A. Bright, P. R. Coleman, and B. L. Bhaduri. 2003. "LandScan: A Global Population Database for Estimating Populations at Risk in Remotely Sensed Cities." In *Remotely Sensed Cities*, edited by V. Mesev, 267–281. London: Taylor and Francis.
- EROS Data Center. 2000. "Global Land Cover Characteristics Database, version 2.0." Accessed on August 1, 2013. http://edc2.usgs.gov/glcc/globdoc2_0.php. Anonymous ftp access at edcftp.cr.usgs.gov.
- FAO. 2001. *Global Forest Resources Assessment 2000*. Forestry Paper 140. Rome: Food and Agriculture Organization of United Nations.
- Fischer, G., F. O. Nachtergaele, S. Prieler, E. Teixeira, G. Tóth, H. van Velthuisen, L. Verelst, and D. Wiberg. 2010. Global Agro-Ecological Zones (GAEZ v3.0). Data set. Accessed October 2013, 2013. <http://gaez.fao.org/Main.html#>.
- Foley, J. A., R. Defries, G. P. Asner, C. Barford, G. Bonan, S. R. Carpenter, F. S. Chapin, M. T. Coe, G. C. Daily, H. K. Gibbs, J. H. Helkowski, T. Holloway, E. A. Howard, C. J. Kucharik, C. Monfreda, J. A. Patz, I. C. Prentice, N. Ramankutty, P. K. Snyder. 2005. Global Consequences of Land Use. *Science* 309 (5734): 570–574. doi: 10.1126/science.1111772.
- Foley, J. A., N. Ramankutty, K. Brauman, E. Cassidy, J. Gerber, M. Johnston, N. Mueller, C. O'Connell, D. K. Ray, P. C. West, C. Balzer, E. M. Bennett, S. R. Carpenter, J. Hill, C. Monfreda, S. Polasky, J. Rockström, J. Sheehan, S. Siebert, D. Tilman, and D. P. M. Zaks. 2011. "Solutions for a Cultivated Planet." *Nature* 478: 337–342. doi: 10.1038/nature10452.
- Friedl, M. A., D. K. McIver, J. C. F. Hodges, X. Y. Zhang, D. Muchoney, A. H. Strahler, C. E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C. Schaaf. 2002. "Global Land Cover Mapping from MODIS: Algorithms and Early Results." *Remote Sensing of Environment* 83: 287–302.
- Golan, A., G. Judge, and D. Miller. 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. New York: Wiley.
- IFPRI (International Food Policy Research Institute). 2002. Global Agricultural Extent v 2.0. Reinterpretation of Global Land Cover Characteristics Database (GLCCD v. 2.0). www.asb.cgiar.org/BNPP/phase2/bnpp. Phase 2 datasets.
- Leff, B., N. Ramankutty, and J. A. Foley. 2004. "Geographic Distribution of Major Crops across the World." *Global Biogeochemical Cycles* 18 (1): n/a–n/a. doi: 10.1029/2003GB002108.
- Lencer, 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.
- Liu, J., L. You, M. Amini, M. Obersteiner, M. Herrero, A. Zehnder, and H. Yang. 2010. "A High-Resolution Assessment of Global Nitrogen Flows in Cropland." *Proceedings of National Academy of Sciences of USA* 107: 8035–8040. doi: 10.1073/pnas.0913658107.

- Monfreda, C., N. Ramankutty, and J. A. Foley. 2008. "Farming the Planet: 2. Geographic Distribution of Crop Areas, Yields, Physiological Types, and Net Primary Production in the Year 2000." *Global Biogeochemical Cycles* 22 (1): n/a–n/a. doi: 10.1029/2007GB002947.
- Portmann, F., S. Siebert, C. Bauer, and P. Döll. 2008. *Global Data Set of Monthly Growing Areas of 26 Irrigated Crops*. Frankfurt Hydrology Paper 06, Institute of Physical Geography, University of Frankfurt, Frankfurt am Main, Germany.
- Portmann, F. T., S. Siebert, and P. Döll. 2010. "MIRCA2000— Global Monthly Irrigated and Rainfed Crop Areas around the Year 2000: A New High-Resolution Data Set for Agricultural and Hydrological Modeling." *Global Biogeochemical Cycles* 24 (1): n/a–n/a. doi: 10.1029/2008GB003435.
- Ramankutty, N., C. Delire, and P. Snyder. 2006. "Feedbacks between Agriculture and Climate: An Illustration of the Potential Unintended Consequences of Human Land Use Activities." *Global and Planetary Change* 54: 19–93. doi: 10.1016/j.gloplacha.2005.10.005.
- Ramankutty, N., A. T. Evan, C. Monfreda, and J. A. Foley. 2008. "Farming the Planet: 1. Geographic Distribution of Global Agricultural Lands in the Year 2000." *Global Biogeochemical Cycles* 22 (1): n/a–n/a. doi: 10.1029/2007GB002952.
- Rosegrant, M., X. Cai, S. Cline. 2002. *World Water and Food to 2025: Dealing with Scarcity*. Washington, DC: International Food Policy Research Institute.
- Siebert, S., P. Doll, J. Hoogeveen, J.-M. Faures, K. Frenken, and S. Feick 2005, "Development and Validation of the Global Map of Irrigation Areas." *Hydrology and Earth System Sciences* 2: 1299–1327.
- Voldoire, A., B. Eickhout, M. Schaeffer, J. F. Royer, and F. Chauvin. 2007. "Climate Simulation of the Twenty-First Century with Interactive Land-Use Changes." *Climate Dynamic* 29: 177–193.
- WDPA (World Database on Protected Areas). 2009. *WDPA Annual Release 2009*. www.wdpa.org/AnnualRelease.aspx.
- Yang, D. W., S. Kanae, T. Oki, T. Koike, and K. Musiake. 2003. "Global Potential Soil Erosion with Reference to Land Use and Climate Changes." *Hydrological Processes* 17: 2913–2928.
- You, L., and S. Wood. 2006. "An Entropy Approach to Spatial Disaggregation of Agricultural Production." *Agricultural Systems* 90 (1–3): 329–347. doi: 10.1016/j.agsy.2006.01.008.
- You, L., S. Wood, and U. Wood-Sichra. 2009. "Generating Plausible Crop Distribution Maps for Sub-Saharan Africa Using a Spatially Disaggregated Data Fusion and Optimization Approach." *Agricultural Systems* 99 (2–3): 126–140. doi: 10.1016/j.agsy.2008.11.003.
- You, L., S. Wood, U. Wood-Sichra, and W. Wu. 2013. "Generating Global Crop Distribution Maps: From Census to Grid." *Agricultural Systems*. Forthcoming.
- 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.

RECENT IFPRI DISCUSSION PAPERS

For earlier discussion papers, please go to www.ifpri.org/pubs/pubs.htm#dp.
All discussion papers can be downloaded free of charge.

1326. *Population pressure and livelihood dynamics: Panel evidence from Bangladesh*. Shahidur Rashid, Nigussie Tefera, Solomon Lemma, and Mohammed Yunus, 2014.
1325. *Changing sources of growth in Indian Agriculture: Implications for regional priorities for accelerating agricultural growth*. Pratap S. Borthal, Pramod K. Joshi, Digvijay S. Negi, and Shaily Agarwal, 2014.
1324. *Risk and Ambiguity Preferences and the Adoption of New Agricultural Technologies: Evidence from Field Experiments in Rural India*. Patrick S. Ward and Vartika Singh, 2014.
1323. *Land, Assets, and Livelihoods: Gendered Analysis of Evidence from Odisha State in India*. Vivien Savath, Diana Fletschner, Amber Peterman, and Florence Santos, 2014.
1322. *Exploring local perceptions of climate change impact and adaptation in rural Bangladesh*. Peter Davis and Snigdha Ali, 2014.
1321. *Closing the gender asset gap: Learning from value chain development in Africa and Asia*. Agnes R. Quisumbing, Deborah Rubin, Cristina Manfre, Elizabeth Waithanji, Mara van den Bold, Deanna Olney, and Ruth Meinzen-Dick, 2014.
1320. *The political economy of Zambia's recovery: Structural change without transformation?* Danielle Resnick and James Thurlow, 2014.
1319. *The impact of Bolsa Familia on schooling: Girls' advantage increases and older children gain*. Alan de Brauw, Daniel O. Gilligan, John Hoddinott, and Shalini Roy, 2014.
1318. *Exploring strategic priorities for regional agricultural research and development investments in Southern Africa*. Michael Johnson, Samuel Benin, Liangzhi You, Xinshen Diao, Pius Chilonda, and Adam Kennedy, 2014.
1317. *How do intrahousehold dynamics change when assets are transferred to women?: Evidence from BRAC's Challenging the Frontiers of Poverty Reduction—Targeting the Ultra Poor Program in Bangladesh*. Narayan Das, Rabeya Yasmin, Jinnat Ara, Md. Kamruzzaman, Peter Davis, Julia A. Behrman, Shalini Roy, and Agnes R. Quisumbing, 2013.
1316. *Is there an app for that?: The impact of community knowledge workers in Uganda*. Bjorn Van Campenhout, 2013.
1315. *Can integrated agriculture-nutrition programs change gender norms on land and asset ownership: Evidence from Burkina Faso*. Mara van den Bold, Abdoulaye Pedehombga, Marcellin Ouedraogo, Agnes R. Quisumbing, and Deanna Olney, 2013.
1314. *Moving in the right direction?: Maize productivity and fertilizer use and use intensity in Ghana*. Antony Chapoto and Catherine Ragasa, 2013.
1313. *Women's empowerment in agriculture, production diversity, and nutrition: Evidence from Nepal*. Hazel Jean L. Malapit, Suneetha Kadiyala, Agnes R. Quisumbing, Kenda Cunningham, and Parul Tyagi, 2013.
1312. *Don't tell on me: Experimental evidence of asymmetric information in transnational households*. Kate Ambler, 2013.
1311. *Can dairy value-chain projects change gender norms in rural Bangladesh?: Impacts on assets, gender norms, and time use*. Agnes R. Quisumbing, Shalini D. Roy, Jemimah Njuki, Kakuly Tanvin, and Elizabeth Waithanji, 2013.
1310. *Can government-allocated land contribute to food security?: Intrahousehold analysis of West Bengal's microplot allocation program*. Florence Santos, Diana Fletschner, Vivien Savath, and Amber Peterman, 2013.
1309. *The National Health Insurance Scheme in Ghana: Implementation challenges and proposed solutions*. Giselle Gajate-Garrido and Rebecca Owusua, 2013.
1308. *Gender inequalities in ownership and control of land in Africa: Myths versus reality*. Cheryl Doss, Chiara Kovarik, Amber Peterman, Agnes R. Quisumbing, Mara van den Bold, 2013.
1307. *Farmer preferences for drought tolerance in hybrid versus inbred rice: Evidence from Bihar, India*. Patrick S. Ward, David L. Ortega, David J. Spielman, Vartika Singh, 2013.
1306. *Borrowing from the insurer: An empirical analysis of demand and impact of insurance in China*. Yanyan Liu, Kevin Chen, Ruth Hill, and Chengwei Xiao, 2013.

**INTERNATIONAL FOOD POLICY
RESEARCH INSTITUTE**

www.ifpri.org

IFPRI HEADQUARTERS

2033 K Street, NW
Washington, DC 20006-1002 USA
Tel.: +1-202-862-5600
Fax: +1-202-467-4439
Email: ifpri@cgiar.org