Agronomía Colombiana
ISSN: 0120-9965
ISSN: 2357-3732
Universidad Nacional de Colombia, Facultad de Agronomía

Arango-Londoño, David; Ramírez-Villegas, Julián; Barrios-Pérez, Camilo; Bonilla-Findji, Osana; Jarvis, Andy; Uribe, Jorge Mario

Closing yield gaps in Colombian direct seeding rice systems: a stochastic frontier analysis

Agronomía Colombiana, vol. 38, no. 1, 2020, January-April, pp. 110-119
Universidad Nacional de Colombia, Facultad de Agronomía

DOI: https://doi.org/10.15446/agron.colomb.v38n1.79470

Available in: https://www.redalyc.org/articulo.oa?id=180376060012

Closing yield gaps in Colombian direct seeding rice systems: a stochastic frontier analysis

Cierre de brechas de rendimiento en los sistemas colombianos de siembra directa de arroz: un análisis de frontera estocástica

David Arango-Londoño[1,2*], Julián Ramírez-Villegas [2,3,4], Camilo Barrios-Pérez[2,3,6], Osana Bonilla-Findji[2,3], Andy Jarvis[2,3], and Jorge Mario Uribe[5]

ABSTRACT

Rice is one of the most important crops in terms of harvested area and food security both globally and for Colombia. Improvement of technical efficiency levels in rice production in order to close yield gaps in a context in which rice demand increases, natural resources are depleted, and where there are growing expectations about both climate changes and trade agreements is likely the most important challenge that farmers confront. This research assessed the main management factors that limit both rice crop productivity and the likely drivers of non-optimal technical efficiency levels (a proxy for yield gaps). This study focused on both upland and irrigated direct seeding systems across a variety of environments in Colombia. Stochastic frontier models were used to integrate microeconomic theory and empirical regression analysis in conjunction with a large commercial rice production database developed by the Colombian rice growers' federation (Fedearroz). A large variation was found in technical efficiency (from 40 to 95%) levels for both upland and irrigated systems, and major differences were obtained in the limiting factors of the two systems (e.g. seed availability, variety type, market accessibility, fertilizer type, and use rate). This suggests both substantial and varied opportunities for improvements in current technical efficiency levels. Across systems, the correct choice of variety was identified as a common key factor for maximizing yield for a particular environment. For upland systems, optimal choices were F174 and F2000, whereas for irrigated rice F473 was found to produce the highest yield. Additionally, numerical analysis suggests a yield impact of ca. 0.18% for each 1% increase in the nitrogen application rate for upland systems. For irrigated rice, phosphorous rather than nitrogen application rates were found to be more important. Since our analysis is based on farm-scale commercial production data, we argue that once our results are brought to consensus with local extension agents, technicians and agronomists, then management recommendations for closing yield gaps can be used to improve rice productivity.

Key words: empirical models, food security, increasing rice demand, production function, technical efficiency.

RESUMEN

El cultivo de arroz es uno de los más importantes en términos de área cosechada y seguridad alimentaria tanto a nivel global como en Colombia. Mejorar los niveles de eficiencia técnica en la producción de arroz para cerrar las brechas de rendimiento en un contexto en el que la demanda de alimento aumenta, los recursos naturales escasean y las expectativas sobre los impactos potenciales del cambio climático y los tratados de libre comercio crecen es probablemente el desafío más importante que enfrentan los agricultores. Esta investigación evaluó los principales factores de gestión que limitan la productividad del cultivo de arroz, y los posibles impulsores de niveles de eficiencia técnica no óptimos (un proxy de las brechas de rendimiento). El estudio se enfocó en sistemas de siembra directa de tierras altas y de riego en diferentes ambientes en Colombia. Utilizamos modelos de frontera estocástica para integrar la teoría microeconómica y el análisis de regresión empírica junto con una gran base de datos de producción comercial de arroz, desarrollada por la Federación colombiana de productores de arroz (Fedearroz). Se encontró una gran variación en los niveles de eficiencia técnica (del 40 al 95%) para los sistemas de tierras altas y de riego, y se obtuvieron diferencias importantes en los factores limitantes entre los dos sistemas (por ejemplo, disponibilidad de semillas, tipo de variedad, accesibilidad al mercado, tipo de fertilizante, y tasa de uso). Esto sugiere oportunidades sustanciales y variadas para mejorar los niveles actuales de eficiencia técnica. En todos los sistemas, la elección correcta de la variedad se identificó como un factor clave común para maximizar el rendimiento por ambiente. Para los sistemas de tierras altas, las opciones óptimas fueron F174 y F2000, mientras que para el arroz de riego se encontró que F473 era el de mayor rendimiento. Además, el análisis numérico sugiere un impacto en el rendimiento de ca. 0.18% por cada 1% de aumento en la tasa de aplicación de nitrógeno para sistemas de tierras altas. Para el arroz de riego, se encontró que las tasas de aplicación de fósforo en lugar de nitrógeno eran más importantes. Como nuestro análisis se basa en datos de producción comercial a escala de finca, se argumentó que una vez que nuestros resultados llegan a un consenso con los agentes de extensión, técnicos y agrónomos locales, las recomendaciones de gestión para cerrar las brechas de rendimiento se pueden utilizar para mejorar la productividad del arroz.

Palabras clave: modelos empíricos, seguridad alimentaria, aumento de la demanda de arroz, función de producción, eficiencia técnica.

Received for publication: 2 May, 2019. Accepted for publication: 23 March, 2020

Doi: 10.15446/agron.colomb.v38n1.79470

[1] Department of Natural Sciences and Mathematics, Pontificia Universidad Javeriana, Cali (Colombia).
[2] Research Program on Climate Change, Agriculture and Food Security (CCAFS), Consultative Group for International Agricultural Research (CGIAR), Cali (Colombia).
[3] International Center for Tropical Agriculture (CIAT), Cali (Colombia).
[4] School of Earth and Environment, Institute for Climate and Atmospheric Science (ICAS), University of Leeds, Leeds (UK).
[5] School of Economics, Universidad del Valle, Cali (Colombia).
[6] Laboratory of International Agricultural Development Studies, School of Global Agricultural Science, University of Tokyo, Tokyo (Japan).
[*] Corresponding author: david.arango@javerianacali.edu.co

Introduction

The challenges in terms of food security for human societies for the coming decades are enormous. Recent projections of human population growth indicate that global population will reach ca. 9 billion by 2050 (FAO, 2010; UN, 2010). Global food demand will double and at the same time competition for the use of food crops to produce bioenergy will increase (Wheeler and von Braun, 2013; Tilman and Clark, 2014). Such increasing demand for agricultural products will only raise the pressure for natural resources that are vital for agriculture, and this will be compounded by land and water competition of expanding urban centers (Foley *et al*., 2011; West *et al*., 2014). Sustainable improvement of crop productivity by closing yield gaps is, thus, a top priority for agriculture across the developing world (Licker *et al*., 2010; Lipper *et al*., 2014).

In Colombia, rice is a highly important crop for both food security and farmers' incomes, given its high consumption and acceptance rates. It is, therefore, essential for Colombian daily dietary requirements (Khoury *et al*., 2014). Rice is the second annual crop in terms of area harvested with a total area of ca. 685,138 ha (30% area in annual crops) in 2018 in Colombia after maize, but in terms of production value it is the first in the nation (DANE, 2016). Colombia is the second largest rice producer in Latin America and the Caribbean (MADR, 2012). In spite of its importance in Colombia, for many developing countries the average on-farm irrigated rice yields are about 50% below their potential (FAO, 2004; Lobell *et al*., 2009; Licker *et al*., 2010). In humid tropical conditions, modern rice varieties yield between 10-11 t ha^{-1}, but irrigated rice yields are typically in the range 4-6 t ha^{-1} for Colombia (FAO, 2004). As in many other countries, sub-optimal management in conjunction with climate variability is the main cause for yield gaps.

In this research, the technical efficiency (a proxy for yield gaps) and its drivers were estimated for a representative sample of rice farmers from different rice growing regions in Colombia using stochastic frontier models (SFMs). In Colombia, only SFMs have been used to quantify technical efficiencies of coffee production (Perdomo and Mendieta, 2007; Perdomo and Hueth, 2011). A large number of applications of stochastic frontier models exist for rice in Asian countries (Xua and Jeffrey, 1998; Mythili and Shanmugam, 2000; Tian and Wan, 2000; Chang and Wen, 2011). The vast majority of these studies focus on estimating production frontiers through stochastic frontier models, typically for panel and cross-sectional datasets. All these studies include variables that are economically directly related to production efficiency functions, such as harvested area, fertilizer applications, and labor and machinery use. However, recent studies also include household socio-economic conditions (Villano and Fleming, 2004) and/or environmental variables, such as temperature and precipitation (Chang and Wen, 2011).

The objectives of this study were to quantify technical efficiencies in rice production across different production environments in Colombia, identify management factors affecting efficiency, and propose realistic changes in management that allow increasing technical efficiency. A database from a survey of 771 representative rice farms from the three main rice producing regions of Colombia (north, center and east) during the period 2007-2015 was used to fit an SFM for a range of environments. The SFMs were built separately for upland (eastern and northern regions) and irrigated systems (central and northern regions) as both systems have very different management regimes and yields. The SFMs were then used to quantify technical efficiencies and their driving factors, and finally to use a sensitivity analysis to assess the management changes required to increase efficiency levels. The findings were discussed in light of potential strategies to enhance farmer incomes and welfare by helping them to reach optimal efficiency levels.

Materials and methods

Stochastic frontier modeling

The overall framework and key concepts used throughout this work were introduced by the following terms. The ability of a given production unit to maximize its output, given a particular set of inputs, is what we term 'technical efficiency'. The overall assumption in our analytic framework is that those units that have the highest productivity form the 'production frontier' that can be estimated from data. Our approach follows that of Kumbhakar and Lovell (2000).

Let q_i be the amount of rice (kg per harvest) produced by farm *i*, defined by a production function *f* (Eq. 1), which is defined by a vector of *j* variables measured (as fertilize application, man, and machine labor) at the farm (X_i) with their associated coefficients ($\hat{\beta}_i$), and a stochastic component η_i.

$$q_i = f\left(\hat{\beta} X_{i,j}\right) + \eta_i \quad (1)$$

The component η_i is formed by two independent elements (Eq. 2).

$$\eta_i = v_i - u_i \quad (2)$$

In summary,

$$q_i = f\left(x_{1i}, x_{2i}, x_{3i}, .., \hat{\beta}\right) + \eta_i \text{ where } \eta_i = v_i - u_i \quad (3)$$

The first element in Equation 2 v_i is associated with random variations (error) and is symmetrical and independently distributed with mean zero and constant variance $v_i \sim N(0, \sigma_v)$ and can take positive or negative values $v_i \in (-\infty, +\infty)$. The second element, u_i, is the technical inefficiency observed in rice production (q_i) that is asymmetric $u_i > N(0, \sigma_u)$ and greater than zero and independent from v_i.

Given the characteristics of η_i and the need to have unbiased and consistent model parameters, the estimators ($\hat{\beta}$) of the stochastic frontier function must be computed using a maximum likelihood approach (Aigner *et al.*, 1977). The natural logarithm of the likelihood function ($Ln\, f$) is defined by Equation 4.

$$Ln\left(\sigma_s^2, \hat{\beta}\right) = -\frac{n}{2}Ln(2\pi) - \frac{n}{2}Ln(\sigma_s^2) + \sum_{i=1}^{n} Ln\left(1 - \varphi(z_i)\right) - \frac{n}{2\sigma_s^2}\sum_{i=1}^{n}\left[q_i - f\left(\hat{\beta}X_{i,j}\right)\right]^2 \quad (4)$$

where n is the total number of observations (i.e. surveyed farms per system), $\sigma^s{}_2$ is the variance of the model (Eq. 5), and $\varphi(z_i)$ is the cumulative normal standard distribution, in which the parameter γ (Eq. 6) represents the efficiency parameters stemming from both error sources in Eq. 3. When the random effect (v_i) dominates (i.e. $\sigma^u{}_2 \rightarrow 0$ and $\gamma \rightarrow 0$), high efficiency occurs in a group of farmers (i.e., most farmers adequately use their inputs to maximize production). Conversely, when the asymmetric component (u_i) tends towards infinite (i.e. $\sigma^u{}_2 \rightarrow \infty$ and $\gamma \geq 1$), the technical inefficiency is the main source of variation in the model (i.e., most farmers use inputs in a non-optimal way and are hence far from the production frontier).

$$\sigma_s^2 = \sigma_u^2 + \sigma_v^2 \quad (5)$$

$$z_i = \frac{q_i - f\left(\hat{\beta}X_{i,j}\right)}{\sigma_s^2}\sqrt{\frac{\gamma}{1-\gamma}} \quad (6)$$

Finally, the technical efficiency level (TE_i) can be estimated for each rice farm as the ratio between the actual production (q_i) and the maximum achievable production (q^*_i) (Eq. 7).

$$TE_i = \frac{q_i}{q_i^*} = \frac{e^{f\left(\hat{\beta}X_{i,j}\right)+v_i-u_i}}{e^{f\left(\hat{\beta}X_{i,j}\right)+v_i}} = e^{-u_i} \quad (7)$$

TE_i represents the ratio between the attained and the potential production. Potential production is defined as the production when the rice farmer uses all inputs efficiently. TE_i varies in the range 0-1 with high values indicative of high technical efficiency and vice versa. The TE_i values can be used to identify farms and municipalities where efficiency is low and, hence, where interventions may be urgently required to close yield gaps.

Study region

Our study region is Colombia, and specifically all provinces where rice is grown commercially (Fig. 1). The region comprises 11 different provinces distributed across the country and across a range of climatic conditions. Farmers in the Eastern llanos (provinces of Meta and Casanare), the driest region, grow primarily upland rice with relatively low inputs. The central region (provinces of Huila and Tolima) features the largest production quantities. In this region, rice is grown only under irrigated conditions with relatively high inputs and in a variety of elevations and climatic conditions. The northern Caribbean region is formed by the provinces of Cordoba, Sucre, Bolivar, Magdalena, Guajira, Cesar, and Norte de Santander and features both irrigated and upland systems. This region is the largest in extent and likely the most diverse in inputs and farmer socio-economic conditions.

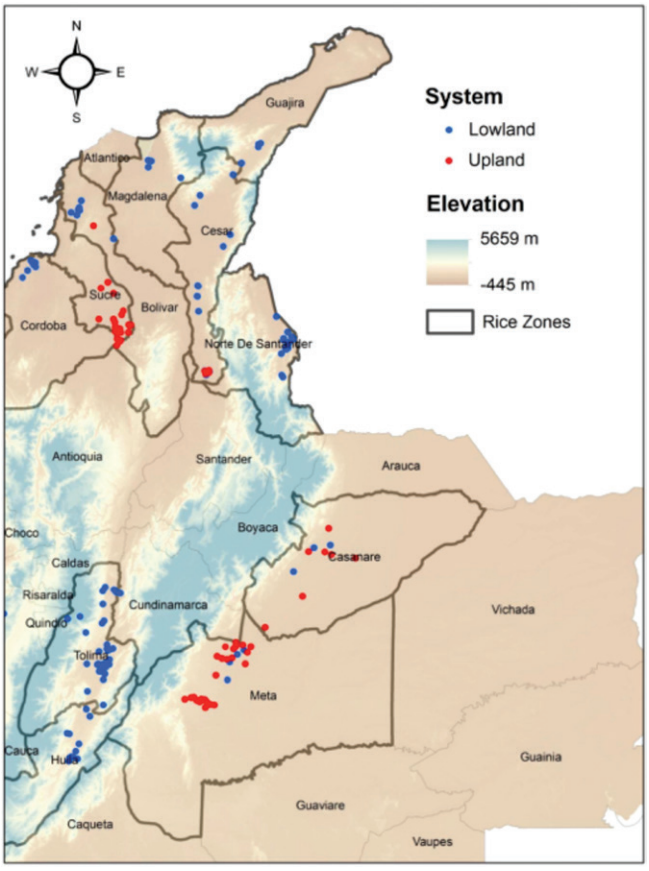

FIGURE 1. Study region and location of rice farms for both irrigated (blue dots) and upland (red dots) systems. Data from Colombia during the period 2007-2015.

Input data

Crop data

Crop management and economic data were gathered from a database of the Colombian National Rice Federation (Fedearroz) through the National Rice Survey (ENA, in Spanish). The survey was carried out in a representative sample of 771 rice farms in the main rice producing areas of Colombia (Fig. 1) for the period 2007-2017, for both irrigated and upland production systems. The survey recorded crop yield, seed quantity, cultivar used, farm size, fertilizer use (quantity and frequency), total use of nitrogen and phosphorous, and man and machine labor hours. As an additional socio-economic factor of potential relevance, we calculated access to farms as the distance from each of the 771 farms to the closest primary, secondary or tertiary road.

Table 1 shows the main characteristics of the two systems. As expected from their contributions to national harvested area and production, there are more irrigated farms (509) than upland farms (262) in the database. There is also a difference between their yield, with irrigated rice yielding ca. 30% more than upland rice (about 1.3 t ha^{-1} more). We note major management differences, with irrigated systems using more fertilizer and more man hours. Due to these differences, we analyzed both systems separately.

TABLE 1. Rice cultivation farm management practices as derived from the National Rice Survey database. Values shown are mean values for each crop system across all observations. Data from Colombia during the period 2007-2015.

Variable / System	Irrigated	Upland	All
Total number of surveyed farms	509	262	771
Yield (kg ha^{-1})	5,992	4,623	5,527
Quantity of seed used (kg ha^{-1})	201.1	203.1	201.8
Average farm size (ha)	54.3	69.1	59.3
Rate of fertilizer per crop cycle (kg ha^{-1})	619.1	401.6	544.9
Total number of fertilizer applications per crop cycle	5.9	3.4	5.1
Rate of total nitrogen use (kg ha^{-1})	154.3	80.9	129.3
Rate of total phosphorous use (kg ha^{-1})	29.9	25.9	28.6
Man labor hours per crop cycle (h)	102.5	77.5	92.5
Machine labor hours per crop cycle (h)	48.8	77.7	60.5

Weather data

We gathered daily weather data for the crop cycle (from sowing to harvest) from the weather station network of the Colombian Institute for Hydrology, Meteorology and Environmental Studies (IDEAM). Total precipitation and minimum and maximum temperature data were available for all weather stations. A quality control process was performed on each of the weather stations using the RClimTool software package (Llanos and Arango, 2015) following Esquivel *et al*. (2018) by detecting and correcting outliers and data gaps. Finally, we assigned the weather information to each farm using geostatistical (autocorrelation) analysis with the geoR library in the R statistical package (R Core Team, 2017). For each farm, we then computed the following variables: total precipitation per crop cycle (P), number of rainy days (i.e. with precipitation above 1 mm, R_D), average minimum temperature for the crop cycle (T_{min}), and average maximum temperature for the crop cycle (T_{max}).

Soil data

Soil data were gathered from the World Soil Information database (ISRIC, 2014). This database consists of globally interpolated chemical and physical soil properties at a resolution of 30 arc-sec (about 1 km around the equator). Using the package 'raster' of the R software version 3.1, we extracted the organic carbon content and the water holding capacity for each of the rice farms.

Environmental classification

As a preliminary step before fitting the SFMs for each crop system (upland, irrigated), climate and soil data were used to derive homogeneous edapho-climatic groups through a cluster analysis using the FactoClass Method (Pardo and Del Campo, 2007). We used a distance index to determine the optimal number of groups. We grouped farms into environments because, regardless of management or socio-economic conditions, the production frontier and optimal achievable production can shift depending upon the environmental conditions prevalent on the farm (Van Wart *et al*., 2013; Heinemann *et al*., 2015). The implication is that the biophysical limit of crop yield varies depending on the prevailing climate (Lobell *et al*., 2009). We thus deemed it necessary to assess these environmental effects through an environmental group classification.

Quantification of technical efficiencies

Once the farms in each production system had been grouped into different environments, we used the 'frontier' package in the R software version 3.1 (Coelli and Henningsen, 2013) to fit the SFMs for each production system. We transformed all continuous variables as well as crop yields into their log forms and used all crop varieties and environmental groups as binary variables in the models. Technical efficiencies for each farm in each crop system were then derived from the models. Finally, we used the model coefficients and the results of an F-test to determine

those variables with strong and statistically significant ($P \leq$ 0.01, $\leq$ 0.05, and 0.1) effects on yield.

Assessing the potential for reducing yield gaps

To determine the needed improvements in crop management that would lead to increased technical efficiencies and, hence, to closing yield gaps, we conducted a sensitivity analysis using the production functions derived for each system. For each system, we performed SFM model runs by modifying the most relevant inputs (as identified by the F-test, above), both separately and in combination, making sure input changes did not fall considerably outside their observed ranges. Specifically, for those inputs having a significant positive impact on yield, we increased their observed values by 5%, 10% and 15%, and for those having a negative impact, we decreased their observed values by 5%, 10% and 15%. All sensitivity runs were performed for each of the rice varieties that had a strong effect on yield and for each environment, so that our results would provide sensible environmental-specific management recommendations for improved rice farming.

Results and discussion

Environmental groups

Clustering results indicate that for upland rice, there are four groups of environments, with largely varying edaphoclimatic characteristics per environment (Tab. 2). We note large differences in total precipitation and in the number of rainy days. Environmental group (EG) 2 received the most precipitation with a relatively poor distribution (53 rainy days), suggesting the occurrence of extremes. This environment also showed the largest differences between maximum and minimum temperatures and soils with high organic carbon contents. Conversely, EG 3 showed the lowest precipitation (1,012 mm per crop cycle), with a more even distribution as compared to EG 2. EG 1 and 4 showed the best temporal distribution of precipitation that may result in yield advantages, although their organic carbon contents were about half those of EG 2.

For irrigated systems, only three EGs were found (Tab. 2). As with the upland systems we found a large variation in the total seasonal precipitation across irrigated EGs. EG 1 had the largest rainfall amounts (904 mm) followed by EG 2 (710 mm), whereas EG 3 showed the lowest seasonal precipitation (450 mm). Large variations were also found for organic carbon, with EG 2 showing the highest content. We noted only small differences between average seasonal maximum and minimum temperatures, suggesting that temperature might not be a limiting factor for irrigated rice yields. The lower (2-3ºC below) maximum temperatures in upland systems compared to irrigated systems highlights the importance of irrigation in allowing evaporative cooling in warm conditions (Peng *et al*., 2004; Julia and Dingkuhn, 2013).

TABLE 2. Environmental groups and their average climatic and soil characteristics for both irrigated and upland rice production. Data from Colombia during the period 2007-2015.

System	EG[1]	P (mm)[1]	R_D (d)[1]	T_{min} (ºC)[1]	T_{max} (ºC)[1]	ASW (%)[1]	OC (%)[1]
Upland	1	1,425	68	18	35	12	13
	2	2,097	53	12	34	12	22
	3	1,012	41	10	35	12	20
	4	1,507	69	20	35	10	11
Irrigated	1	904	52	17	36	12	13
	2	710	47	17	36	11	29
	3	450	31	16	37	10	10

[1]EG: environmental group, P: total precipitation per crop cycle, R_D: number of rainy days (i.e. with precipitation above 1 mm), T_{min}: average minimum temperature for the crop cycle, T_{max}: average maximum temperature for the crop cycle, ASW: available soil water, OC: soil organic carbon content.

Production frontiers and technical efficiencies

Upland rice

For upland rice, a statistically significant variable was the nitrogen application rate, with a yield impact of ca. 0.18% for each 1% increase in this rate (Tab. 3). This result is expected (Nagai and Makino, 2009; Mueller *et al*., 2012), and is also consistent with previously reported nitrogen yield impacts of 0.08% in a study for Taiwan (Chang and Wen, 2011). The amount of seed was also found to have a significant impact on yield, with a regression coefficient of 0.08 ± 0.04. This result can be explained by the fact that Colombian rice systems are directly seeded and large amounts of seed are often required to ensure appropriate germination rates and sufficiently high yields.

Contrary to what would be expected, we found a negative association (regression coefficient of -0.1 ± 0.04) between man labor hours and yield. While this is somewhat counterintuitive, further analysis of the database indicated that most man labor hours are spent on harvesting activities that would reduce the overall harvesting time but would not increase productivity. It is thus important that increases in man labor time are targeted towards activities that have a greater impact on yield such as fertilizer applications, sowing, and pests, diseases, and weed control.

The survey included seven different varieties, with Fedearroz 733 (F733) being the most common in irrigated systems

and overall in the country and F473 being the most used in upland systems, but also to some extent in irrigated systems (Fig. 2). Other varieties such as F2000, F174 and F60 are less often used, although some of them (particularly F2000) have been commercially released only recently. We find that most terms associated with varietal choices were also negatively associated with yields. Our models indicate that F2000 and F473 have no significant difference from the base category (F174), although F473 tends to yield less and F2000 tends to yield slightly more than F174 (Fig. 3B). The other varieties (F733, Fortaleza and IA550) all show statistically significant and negative effects on yields (-0.23, -0.27, and -0.18, respectively, all at $P \leq 0.01$). Since coefficients associated with these varieties were the largest in magnitude across all variables, we conclude that varietal choice is the most important management factor in upland rice systems, with optimal choices being F174 and F2000.

The rest of the variables showed no significant impact, but this does not necessarily mean that they do not have impact under specific conditions or farms. For instance, pesticide doses are not important if the variety used is pest-resistant, but it would be important if the selected variety were sensitive to pest attacks. High noise in the recorded data or low variability across farms could also have precluded the identification of a relevant variable for the yield response. However, for this farm population there are no statistically significant responses for these variables.

Finally, we estimated technical efficiencies for each farmer and found an average technical efficiency level of 77% across all upland rice farms. Despite this relatively high value, we observed some farmers with relatively low

TABLE 3. SFA Coefficient estimates for upland rice systems. Data from Colombia during the period 2007-2015.

Variable	Coefficient	Standard Error	Z Statistic	P Value
(Intercept)	7.39	1.89	3.91	0.00 ***
log(nitrogen)	0.18	0.10	1.88	0.06 *
log(phosphorous)	-0.03	0.04	-0.73	0.46
log(quantity of seed)	0.08	0.04	1.82	0.07 *
log(crop area)	0.01	0.02	0.46	0.64
log(fertilizer use rate)	0.04	0.07	0.60	0.55
log(fertilizer application)	-0.12	0.08	-1.49	0.14
log(pesticide use rate)	-0.06	0.05	-1.13	0.26
log(pesticide application)	-0.01	0.07	-0.15	0.88
log(access)	0.08	0.16	0.51	0.61
log(man hours of labour)	-0.10	0.04	-2.57	0.01 **
log(machine hours of labour)	0.00	0.03	-0.02	0.99
Variety Fedearroz _ 2000	0.01	0.08	0.08	0.94
Variety Fedearroz _ 473	-0.16	0.17	-0.97	0.33
Variety Fedearroz _ 733	-0.23	0.07	-3.33	0.00 ***
Variety Fortaleza	-0.27	0.05	-5.34	0.00 ***
Variety Improarroz _ 1550	-0.18	0.06	-2.81	0.00 ***
Environment 2	-0.10	0.17	-0.58	0.56
Environment 3	-0.13	0.09	-1.49	0.14
Environment 4	0.06	0.07	0.79	0.43
	0.13	0.03	5.25	0.00 ***
	0.95	0.04	21.93	0.00 ***

*** Significant at 1% ** Significant at 5% * Significant at 10%

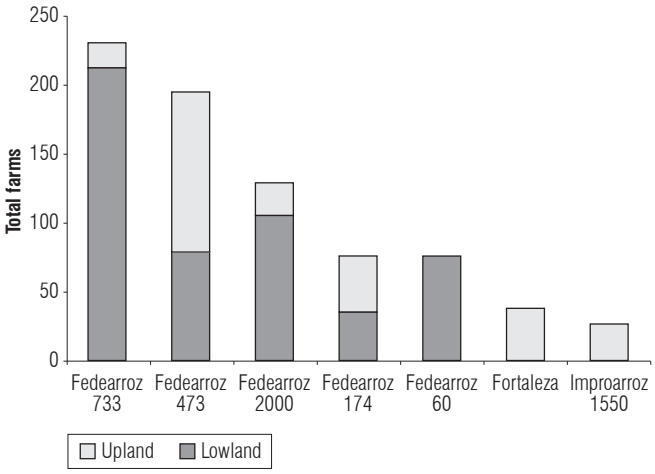

FIGURE 2. Frequency of use of different rice varieties for both rice production systems, as derived from the database used in this study. Data from Colombia during the period 2007-2015.

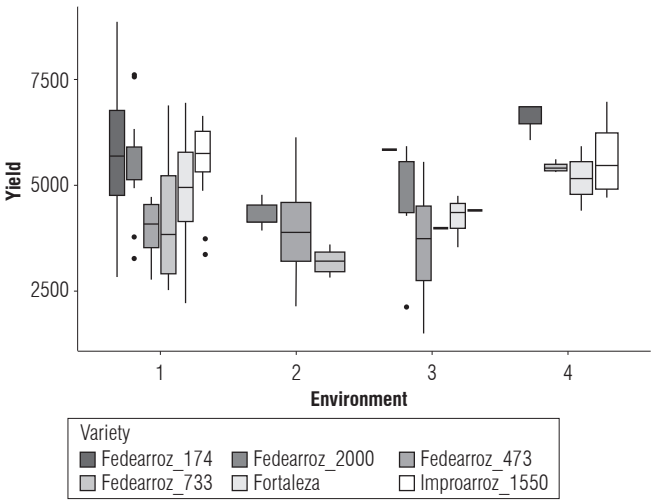

FIGURE 3. Effects of management variables and environmental groups on rice yields for upland rice systems. (A) Model coefficients and (B) yield of all rice varieties for each environment. Horizontal lines show the median, boxes show the interquartile range, whiskers extend to 5-95% of the data, and dots are outliers. Data from Colombia during the period 2007-2015.

efficiencies (below 60%), particularly in the municipalities of Villavicencio, Cumaral, Puerto Lopez and Fuente de Oro, all of which are located in eastern Colombia (Fig. 4).

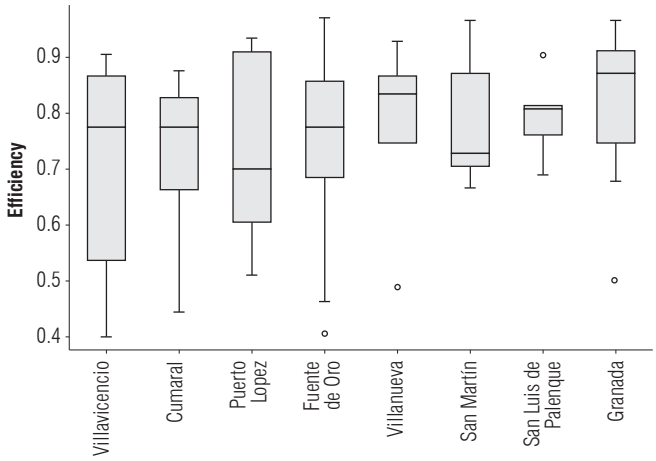

FIGURE 4. Technical efficiencies of upland rice farmers in all municipalities where upland rice is grown. Horizontal lines show the median, boxes show the interquartile range, whiskers extend to 5-95% of the data, and dots are outliers. Data from Colombia during the period 2007-2015.

Irrigated rice

Relevant variables in irrigated systems are different to those in upland systems. This is confirmed by developing system-specific models (Tab. 4). The phosphorous application rate was amongst the most important variables for irrigated rice (coefficient of 0.04). Its positive effect on yield as well as the fact that it had a smaller effect than the nitrogen application rate is in broad agreement with theory (Longstreth and Nobel, 1980; Fujisaka *et al*., 1994). The number of fertilizer applications was also found to be statistically significant, suggesting that a larger number of better temporally distributed yet lower-dosage applications ensures such applications are more efficiently used by the crop in contrast to a reduced number of higher-rate applications.

Access to farms, calculated as the distance from each farm to primary, secondary and tertiary roads was found to affect yields significantly and negatively (coefficient of -0.12). This negative effect indicates that farms that are further away from roads have lower yields. Large distances from farm to roads can limit the timely access of fertilizers and machinery. This could negatively impact farm activity schedules and cause fertilizer, pesticide, or herbicide applications, or machinery-related activities to happen at non-optimal times, thus reducing yields.

Finally, both environments and varieties had a significant impact on yield. In particular, we found that both EG2 and EG3 had a yield advantage as compared to EG1 and that

TABLE 4. Stochastic Frontier Analysis coefficients estimate for irrigated rice system. Data from Colombia during the period 2007-2015.

Variable	Coefficients	Standard Error	Z Statistics	P Value
(Intercept)	9.14	0.55	16.66	0.00 ***
log(nitrogen)	0.05	0.04	1.18	0.24
log(phosphorous)	0.04	0.01	3.18	0.00 ***
log(quantity of seeds)	0.03	0.04	0.63	0.53
log(crop area)	-0.01	0.01	-1.33	0.18
log(fertilizer use rate)	0.07	0.06	1.18	0.24
log(fertilizer aplication)	0.07	0.04	1.66	0.09 *
log(pesticide use rate)	-0.02	0.03	-0.60	0.55
log(pestiside aplication)	-0.04	0.03	-1.23	0.22
log(access)	-0.12	0.03	-3.56	0.00 ***
log(man hours of labour)	0.02	0.04	0.41	0.68
log(machine hours of labour)	0.00	0.02	0.11	0.91
Variety Fedearroz2000	-0.06	0.05	-1.13	0.26
Variety Fedearroz2000473	0.10	0.06	1.63	0.10 *
Variety Fedearroz200060	-0.02	0.06	-0.31	0.76
Variety Fedearroz2000733	-0.06	0.05	-1.09	0.28
Environment 2	0.08	0.04	2.23	0.03 **
Environment 3	0.15	0.03	4.57	0.00 ***
	0.11	0.01	8.23	0.00 ***
	0.97	0.02	50.14	0.00 ***

*** Significant at 1% ** Significant at 5% * Significant at 10%

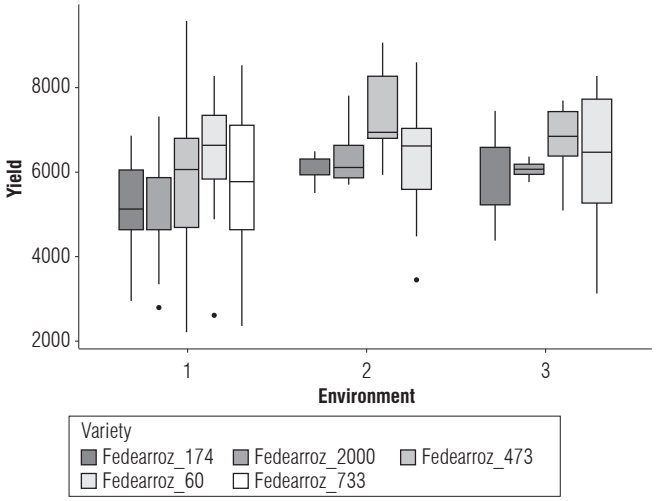

FIGURE 5. Effects of management variables and environmental groups on rice yields for irrigated rice systems. (A) Model coefficients; and (B) yield of all rice varieties for each environment. Horizontal lines show the median, boxes show the interquartile range, whiskers extend to 5-95% of the data, and dots are outliers. Data from Colombia during the period 2007-2015.

the variety F473 had a significantly positive impact (coefficient of 0.1) compared to F174. The other varieties had negative albeit not statistically significant yield impacts (Fig. 5). Our analysis suggests that adequately selecting varieties per environment is critical to increasing yields in irrigated systems (Fig. 5B). The analyses generally highlight the importance of including site-specific variations in yield response models so as to allow the identification of both constraints and opportunities at the scales that are relevant for farmers and extension agents (Jiménez *et al*., 2009, 2011; Delerce *et al*., 2016).

Technical efficiencies for irrigated rice are shown in Figure 6. Notably, the average technical efficiency was 78%, which is very similar to the average efficiency of the upland rice system. However, there are large variations across farmers and environments. Municipalities such as Lorica, La Jagua and Cucuta show low efficiency levels and hence large yield gaps (>40%) for some farmers, whereas the municipalities of Espinal, San Martin, Valledupar and Purificacion show consistently high efficiency levels across farmers. This is consistent with evidence that farmers in northern Colombia are less technologically developed (DANE, 2004).

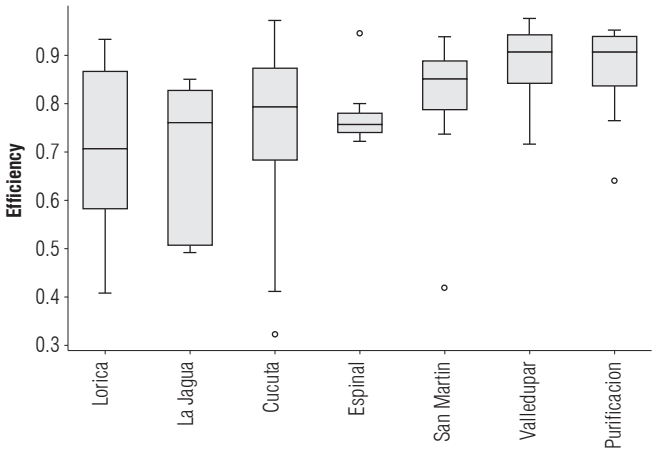

FIGURE 6. Technical efficiencies of irrigated rice farmers in all municipalities where irrigated rice is cultivated. Horizontal lines show the median, boxes show the interquartile range, whiskers extend to 5-95% of the data, and dots are outliers. Data from Colombia during the period 2007-2015.

Conclusions

The stochastic production frontier methodology is a useful alternative to more classical yield gap assessments based on detailed process-based modeling (Bhatia *et al*., 2008; van Bussel *et al*., 2015), and has two clear advantages. First, the method integrates microeconomic theory and empirical analysis, thus allowing the empirical validation of both biophysical and economic hypotheses regarding the cropping systems under study. And secondly, since the method is based on commercial farming information, it has the potential to provide recommendations that are tailored to the farms and farm systems analyzed. The use of georeferenced information at the farm level from a geographically widespread sample of rice farms allowed us to integrate soil and climate data. This not only helped to reduce noise in the responses to environment factors but it also enhanced the capabilities of the methodology beyond only identifying management factors, identifying management x environment interactions, and allowing a more comprehensive assessment of the rice crop systems.

For the upland rice system, our analysis suggests that increasing nitrogen fertilization and seed quantity in upland systems is needed for yield gap closure. Conversely, the amount of man labor hours was found to have a negative effect on productivity, which we attribute to non-optimal use of an excessively large number of man labor hours on harvesting activities. For irrigated systems, we identified the phosphorous application rate as well as the number of fertilizer applications as having a significant and positive effect on yield. For irrigated rice, we also report that accessibility is a limiting factor for a number of farms.

For both rice production systems, we identify two critical aspects that need to be well-managed so as to ensure high levels of production. Most critical is the correct matching of varieties to the prevailing climates of the farms. Our analysis indicates significantly different effects from varieties for the different environments for both production systems. For municipalities with low average efficiency levels, we suggest that improved technical assistance through extension services will be needed to ensure optimal management. For upland systems, we identified Villavicencio, Cumaral, Puerto Lopez and Fuente de Oro, whereas for irrigated systems it would be Lorica, La Jagua and Cucuta. Such technical assistance should be targeted at providing site-specific recommendations and could be based on both our results and the experience of the local extension agents and farmers. Finally, we note that, as this study is based on commercial farming practices and yields across all rice growing areas of Colombia, these recommendations have the potential to directly benefit rice farmers.

Acknowledgments

The authors would like to thank the National Rice Growers Federation (Fedearroz) for providing the crop data that made this study possible. We also thank the Colombian Institute for Hydrology, Meteorology and Environmental

Studies (IDEAM) for recording, storing and providing weather station data used in this study.

Literature cited

Aigner, D., C.A.K. Lovell, and P. Schmidt. 1977. Formulation and estimation of stochastic frontier production function models. J. Econometrics 6, 21-37. Doi: 10.1016/0304-4076(77)90052-5

Bhatia, V.S., P. Singh, S.P. Wani, G.S. Chauhan, A.V.R.K. Rao, A.K. Mishra, and K. Srinivas. 2008. Analysis of potential yields and yield gaps of rainfed soybean in India using CROPGRO-Soybean model. Agric. For. Meteorol. 148, 1252-1265. Doi: 10.1016/j.agrformet.2008.03.004

Chang, H.H. and F.I. Wen. 2011. Off-farm work, technical efficiency, and rice production risk in Taiwan. Agric. Econ. 42, 269-278. Doi: 10.1111/j.1574-0862.2010.00513.x

Coelli, T. and A. Henningsen. 2013. Frontier: stochastic frontier analysis. R package version 1.1-0. URL: https://CRAN.R-Project.org/package=frontier (accessed July 2018).

DANE. 2004. Encuesta Nacional Agropecuaria. URL: https://www.dane.gov.co/files/investigaciones/boletines/ena/Bolet_Arroz_2004.pdf (accessed July 2018).

DANE. 2016. Statistical database. URL: https://www.dane.gov.co/files/investigaciones/agropecuario/censo-nacional-arrocero/boletin-tecnico-4to-censo-nacional-arrocero-2016.pdf (accessed July 2018).

Delerce, S., H. Dorado, A. Grillon, M.C. Rebolledo, S.D. Prager, V.H. Patiño, G. Garcés-Varón, and D. Jiménez. 2016. Assessing weather-yield relationships in rice at local scale using data mining approaches. PLOS One 11(89), e0161620. Doi: 10.1371/journal.pone.0161620

Esquivel, A., L. Llanos-Herrera, D. Agudelo, S.D. Prager, K. Fernandes, A. Rojas, J.J. Valencia, and J. Ramirez-Villegas. 2018. Predictability of seasonal precipitation across major crop growing areas in Colombia. Clim. Serv. 12, 36-47. Doi: 10.1016/j.cliser.2018.09.001

FAO. 2004. Rice and yield gap reduction, International rice year. Rome. URL: http://www.fao.org/3/Y5167E/y5167e02.htm (accessed July 2018).

FAO. 2010. The state of food insecurity in the world. Rome. URL: http://www.fao.org/publications/sofi/2010/en/ (accessed July 2018).

Foley, J.A., N. Ramankutty, K.A. Brauman, E.S. Cassidy, J.S. Gerber, M. Johnston, N.D. 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. Rockstrom, 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

Fujisaka, S., L. Harrington, and P. Hobbs. 1994. Rice-Wheat in South Asia: systems and long-term priorities established through diagnostic research. Agric. Syst. 46, 169-187. Doi: 10.1016/0308-521X(94)90096-X

Heinemann, A.B., C. Barrios-Pérez, J. Ramírez-Villegas, D. Arango-Londoño, O. Bonilla-Findji, J.C. Medeiros, and A. Jarvis. 2015. Variation and impact of drought-stress patterns across upland rice target population of environments in Brazil. J. Exp. Bot. Doi: 10.1093/jxb/erv126

ISRIC. 2014. World soil Information database. URL: https://soilgrids.org/#!/?layer=ORCDRC_M_sl2_250m&vector=1 (accessed August 2018).

Jiménez, D., J. Cock, H.F. Satizábal, M.A. Barreto, A.A. Pérez-Uribe, A. Jarvis, and P. Van Damme. 2009. Analysis of Andean blackberry (*Rubus glaucus*) production models obtained by means of artificial neural networks exploiting information collected by small-scale growers in Colombia and publicly available meteorological data. Comput. Electron. Agric. 69, 198-208. Doi: 10.1016/j.compag.2009.08.008

Jiménez, D., J. Cock, A. Jarvis, J. Garcia, H.F. Satizábal, P. Van Damme, A. Pérez-Uribe, and M.A. Barreto-Sanz. 2011. Interpretation of commercial production information: a case study of lulo (*Solanum quitoense*), an under-researched Andean fruit. Agric. Syst. 104, 258-270. Doi: 10.1016/j.agsy.2010.10.004

Julia, C. and M. Dingkuhn. 2013. Predicting temperature induced sterility of rice spikelets requires simulation of crop-generated microclimate. Eur. J. Agron. 49, 50-60. Doi: 10.1016/j.eja.2013.03.006

Khoury, C.K., A.D. Bjorkman, H. Dempewolf, J. Ramirez-Villegas, L. Guarino, A. Jarvis, L.H. Rieseberg, and P.C. Struik. 2014. Increasing homogeneity in global food supplies and the implications for food security. Proc. Natl. Acad. Sci. U.S.A. 111, 4001-4006. Doi: 10.1073/pnas.1313490111

Kumbhakar, S.C. and K. Lovell. 2000. Stochastic frontier analysis. Cambridge Univeristy Press, Cambridge, UK. Doi: 10.1111/1467-8276.t01-1-00317

Licker, R., M. Johnston, J.A. Foley, C. Barford, C.J. Kucharik, C. Monfreda, and N. Ramankutty. 2010. Mind the gap: how do climate and agricultural management explain the "yield gap" of croplands around the world? Glob. Ecol. Biogeogr. 19, 769-782. Doi: 10.1111/j.1466-8238.2010.00563.x

Lipper, L., P. Thornton, B.M. Campbell, T. Baedeker, A. Braimoh, M. Bwalya, P. Caron, A. Cattaneo, D. Garrity, K. Henry, R. Hottle, L. Jackson, A. Jarvis, F. Kossam, W. Mann, N. McCarthy, A. Meybeck, H. Neufeldt, T. Remington, P.T. Sen, R. Sessa, R. Shula, A. Tibu, and E.F. Torquebiau. 2014. Climate-smart agriculture for food security. Nat. Clim. Chang. 4, 1068-1072. Doi: 10.1038/nclimate2437

Lobell, D.B., K.G. Cassman, and C.B. Field. 2009. Crop yield gaps: their importance, magnitudes, and causes. Annu. Rev. Environ. Resour. 34, 179-204. Doi: 10.1146/annurev.environ.041008.093740

Longstreth, D.J. and P.S. Nobel. 1980. Nutrient influences on leaf photosynthesis. Plant Physiol. 65, 541-543. Doi: 10.1104/pp.65.3.541

Llanos, L. and D. Arango. 2015. RClimTool: a free application for analyzing climatic series. Working Paper International Center for Tropical Agriculture. URL: https://cgspace.cgiar.org/handle/10568/63482 (accessed August 2018).

MADR. 2012. Anuario estadístico del sector agropecuario 2012. URL: https://www.agronet.gov.co/Noticias/Paginas/Noticia842.aspx (accessed July 2018).

Mueller, N.D., J.S. Gerber, M. Johnston, D.K. Ray, N. Ramankutty, and J.A. Foley. 2012. Closing yield gaps through nutrient and water management. Nature 490, 254-257. Doi: 10.1038/nature11420

Mythili, G. and K.R. Shanmugam. 2000. Technical efficiency of rice growers in Tamil Nadu: a study based on panel data. Indian J. Agric. Econ. 55, 15-25. Doi: 10.22004/ag.econ.297715

Nagai, T. and A. Makino. 2009. Differences between rice and wheat in temperature responses of photosynthesis and plant growth. Plant Cell Physiol. 50, 744-55. Doi: 10.1093/pcp/pcp029

Pardo, C.E. and C. Del Campo. 2007. Combinación de métodos factoriales y de análisis de conglomerados en R: el paquete FactoClass. Rev. Colomb. Estad. 30, 231-245.

Peng, S., J. Huang, J.E. Sheehy, R.C. Laza, R.M. Visperas, X. Zhong, G.S. Centeno, G.S. Khush, and K.G. Cassman. 2004. Rice yields decline with higher night temperature from global warming. Proc. Natl. Acad. Sci. U.S.A. 101, 9971-9975. Doi: 10.1073/pnas.0403720101

Perdomo, J. and J. Mendieta. 2007. Factores que afectan la eficiencia técnica y asignativa en el sector cafetero colombiano: una aplicación con análisis envolvente de datos. Desarrollo y Sociedad 2007-II, 1-45.

Perdomo, J. and D. Hueth. 2011. Funciones de producción y eficiencia técnica en el eje cafetero colombiano: una aproximación con frontera estocástica. Rev. Colomb. Estad. 34, 377-402. Doi: 10.22004/ag.econ.100873

R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Tian, W. and G.H. Wan. 2000. Technical Efficiency and Its Determinants in China's Grain Production. J. Product. Anal. 13, 159-174. Doi: 10.1023/A:1007805015716

Tilman, D. and M. Clark. 2014. Global diets link environmental sustainability and human health. Nature 515, 518-522. Doi: 10.1038/nature13959

UN. 2010. World Population Prospects: The 2010 Revision. URL: https://www.oecd-ilibrary.org/content/publication/08b807d4-en (accessed July 2018).

Van Bussel, L.G.J., P. Grassini, J. Van Wart, J. Wolf, L. Claessens, H. Yang, H. Boogaard, H. de Groot, K. Saito, K.G. Cassman, and M.K. van Ittersum. 2015. From field to atlas: upscaling of location-specific yield gap estimates. F. Crop. Res. 177, 98-108. Doi: 10.1016/j.fcr.2015.03.005

Van Wart, J., L.G.J. van Bussel, J. Wolf, R. Licker, P. Grassini, A. Nelson, H. Boogaard, J. Gerber, N.D. Mueller, L. Claessens, M.K. van Ittersum, and K.G. Cassman. 2013. Use of agro-climatic zones to upscale simulated crop yield potential. F. Crop. Res. 143, 44-55. Doi: 10.1016/j.fcr.2012.11.023

Villano, R. and E. Fleming. 2004. Analysis of technical efficiency in rainfed lowland rice environment in Central Luzon Philippines using stochastic frontier production function with heteroskedastic error structure. Working paper series in Agricultural and Resource Economics. University of New England, Armidale, Australia. Doi: 10.22004/ag.econ.12906

West, P.C., J.S. Gerber, P.M. Engstrom, N.D. Mueller, K.A. Brauman, K.M. Carlson, E.S. Cassidy, M. Johnston, G.K. MacDonald, D.K. Ray, and S. Siebert. 2014. Leverage points for improving global food security and the environment. Science 345, 325-328. Doi: 10.1126/science.1246067

Wheeler, T. and J. von Braun. 2013. Climate change impacts on global food security. Science 341(6145), 508-513. Doi: 10.1126/science.1239402

Xua, X. and S.R. Jeffrey. 1998. Efficiency and technical progress in traditional and modern agriculture: evidence from rice production in China. Agric. Econ. 18, 157-165. Doi: 10.1111/j.1574-0862.1998.tb00495.x