

El agro colombiano en píxeles: nuevos aportes a la relación inversa tamaño-productividad

Camilo Erasso

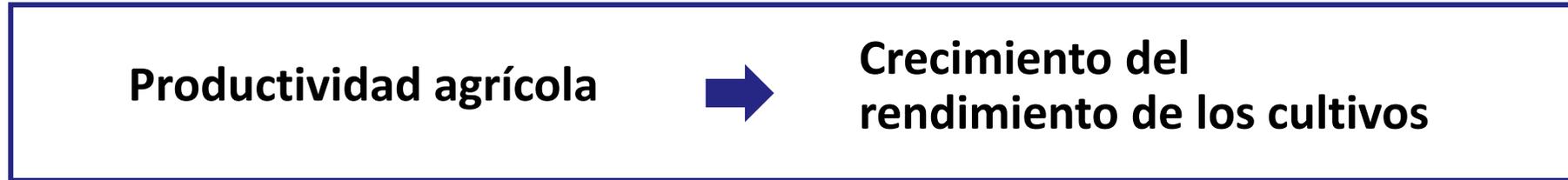
Asistente de Investigación CESED - Uniandes

cf.erasso@correouniandes.edu.co

17 de diciembre de 2020

1. Introducción

Desarrollo Económico



Principal mecanismo para reducir la pobreza
(Carletto, Jolliffe, & Banerjee, 2015)

- Mejor seguridad alimentaria
 - Mayores ingresos agrícolas
- FAO (2017)

Pero existe una incógnita...

Empíricamente, las parcelas pequeñas producen más por hectárea que las grandes



**Acertijo de la relación inversa entre el tamaño del predio y la
productividad por unidad de tierra**

(Bevis & Barrett, 2020)

1. Introducción

¿Acertijo?

Teoría neoclásica: sin fallas de mercado, la productividad de los factores debe ser igual entre las parcelas



Asignación de recursos ineficiente que justifica una redistribución de la tierra. (Barrett, Bellemare, & Hou, 2010)

De gran interés son los hallazgos de **Bevis & Barrett (2020)**:

- en las áreas rurales de Uganda la relación inversa surge a nivel de **parcela**, y no a nivel de finca;
- esta se explica por una variable construida a partir del **cociente entre el perímetro y el área de la parcela**;
- y este cociente refleja un **efecto de borde** (*edge effect*) agronómico (productividad más alta en la periferia de las parcelas)

Objetivo: con nuevos datos, probar directamente la existencia de este efecto de borde y examinar de forma independiente su posible incidencia sobre la relación tamaño-productividad.

2. Estado del arte

Explicaciones a la relación inversa tamaño-productividad:

1. Fallas en los mercados

- Uso intensivo de mano de obra, y diferencia entre salario y costo real de trabajo (Sen, 1966)
- Costos de supervisión en fincas grandes (Taslim, 1989)
- Ventajas crediticias basadas en el tamaño de la tierra (Feder, 1985)
- Uso intensivo de trabajo como seguro (Barrett, 1996)

Representan heterogeneidades (no observables) entre los hogares que deberían de afectar de igual forma a cada parcela del hogar.

Assuncao & Braido (2007) encontraron lo contrario.

2. Relación estadística espuria

- Sesgos ocasionados por errores de medición en el área de las fincas (Lamb, 2003)
- Omisión de variables relevantes como la calidad del suelo (Benjamin, 1995)

- Carletto, Savastano, & Zezza, (2013) hallan lo contrario con mediciones por GPS
- Barrett et al. (2010) demuestran que esa variable no es relevante

2. Estado del arte

Efecto de borde:

En ecología, es el resultado de la interacción entre dos ecosistemas adyacentes separados por una transición abrupta que tiene lugar en sus bordes (Murcia, 1995).

En agroecología, caracteriza alteraciones en el rendimiento de los cultivos hacia el borde de las parcelas causadas por elementos circundantes como *setos o arbustos, bosques, hierbas o maleza, otros cultivos, vías agrícolas*, etc. (Raatz et al., 2019) o por procesos de colonización de *plagas o patógenos*. (Arruda et al., 2020).

Mecanismos biofísicos: **competencia** por el agua y nutrientes a nivel de raíces, o por la luz a nivel de las hojas; **compactación** de suelo, **migración** de plagas por reproducción y alimento.

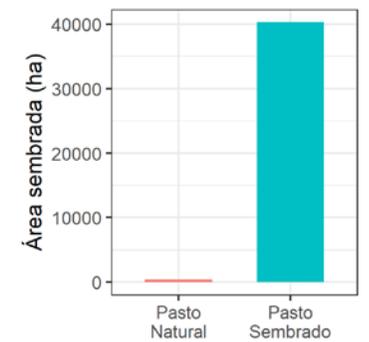
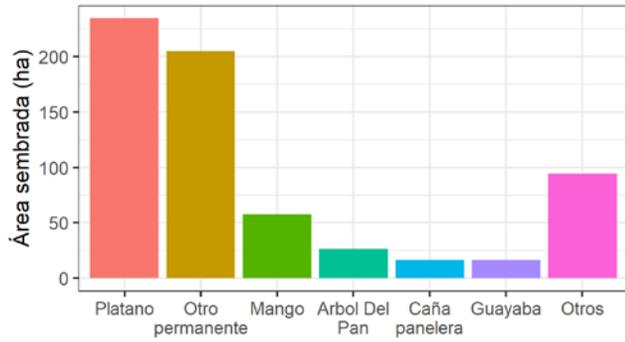
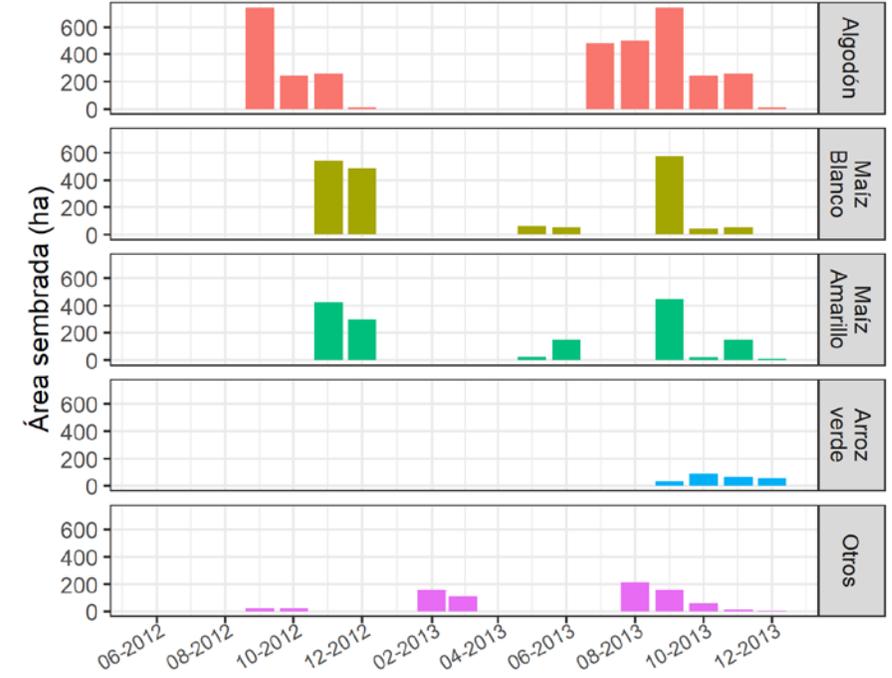
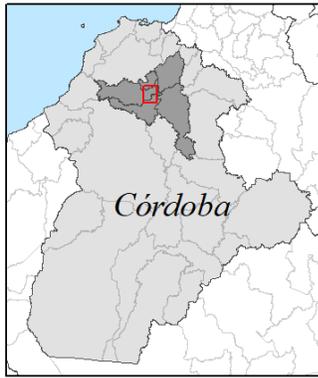
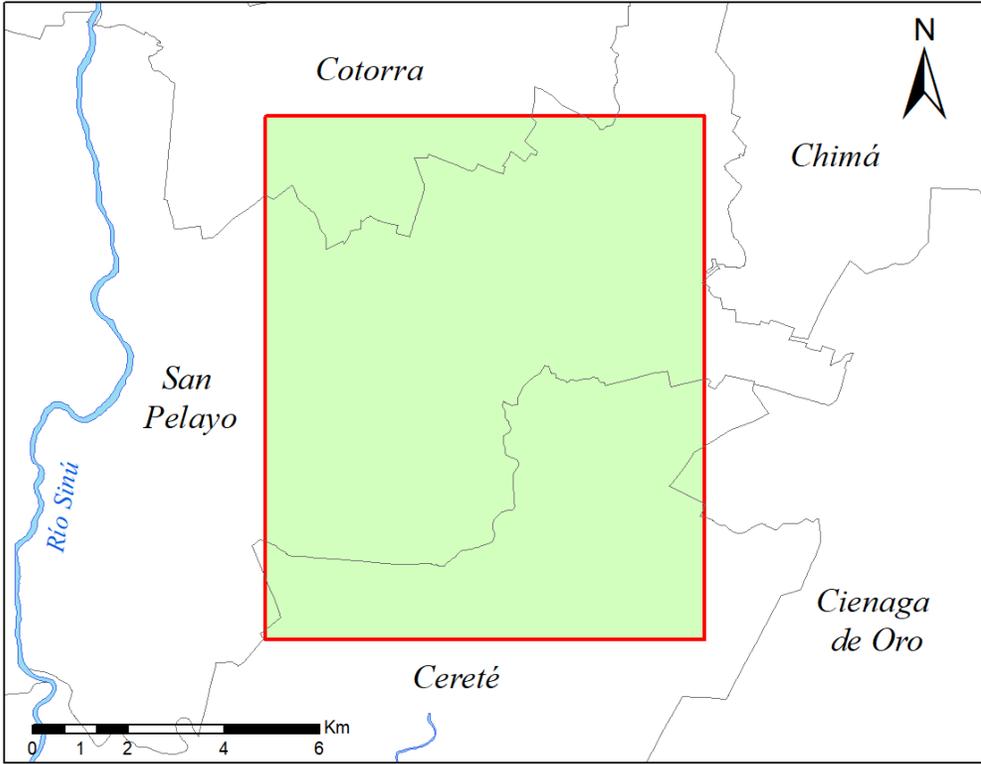
Bevis & Barrett (2020) consideran un mecanismo **comportamental**:

Los campesinos pueden gestionar los bordes de sus parcelas de manera distinta porque son más *accesibles o visibles* en la periferia.

De forma indirecta encuentran que las parcelas con una mayor proporción de área a lo largo de su perímetro tienen una mayor productividad (ej. relación directa).

3. Datos

Región de análisis (≈100 km²)



Fuente: Encuestas georreferenciadas del 3er Censo Nacional Agropecuario (CNA).

3. Datos

Rendimiento de cultivos y el área de parcelas, por:

Encuestas de campo

- Son costosas e imprácticas
- No ofrecen mediciones detalladas a nivel de parcela durante varios periodos de tiempo
- Las cifras autorreportadas están sub y sobrevaloradas (Carletto et al., 2015, 2013)
- Mediciones de área por GPS y de rendimiento de cultivos por *crop cutting* aumentan más los costos y tiempos (Carletto et al., 2015)
- No hay medidas estándar de productividad

Imágenes satelitales

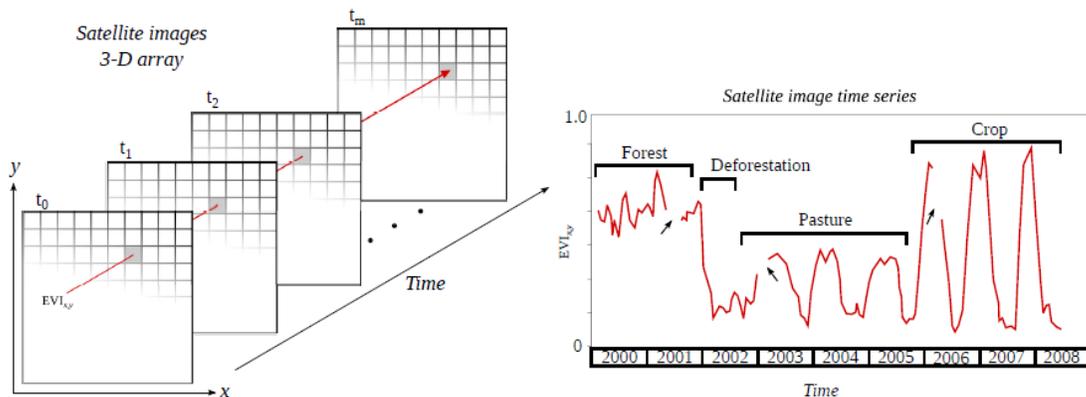
- Nivel más desagregado que la parcela (píxel)
- Ofrecen datos panel a un costo marginal bajo
- Se obtienen proxies de características de la superficie terrestre
- El tamaño del píxel se asemeja al nivel en el que los agricultores toman decisiones (0,5 a 250m)
- Cobertura geográfica bastante amplia (Donaldson & Storeygard, 2016)

Objetivos específico: Construcción de una base de datos **novedosa, escalable y reproducible** a partir de imágenes satelitales con el píxel como unidad de análisis y la parcela como segundo nivel de agregación (ej. **panel píxel-parcela**)

3. Datos

- Satélites
- Tipo de sensores (pasivos o activos)
- Bandas espectrales (RGB + infrarrojo)
- Resolución espacial
- Resolución temporal
- Índice de Vegetación de Diferencia Normalizada (NDVI)

$$NDVI = \frac{NIR - Red}{NIR + Red}$$
- Comportamiento a lo largo del tiempo en el espectro electromagnético



Fuente: Maus, Câmara, Appel, & Pebesma (2017)

	MODIS Terra y Aqua	Landsat 7 y 8	RapidEye
Resolución espacial utilizada	250 m	30 m	5 m
Tipo de producto	Surface Reflectance Daily L2G (M*D09GQ) Version 6	Surface Reflectance Level 2	RapidEye Ortho Level 3A
Bandas espectrales utilizadas	Red, NIR	Red, NIR	Blue, Green, Red, Red Edge, NIR
Res. Temporal (no efectiva)	1 vez al día	Cada 16 días	-
No. Satélites	2	2	5 (constelación de nanosatélites)
Res. Temporal conjunta (no efectiva)	2 veces al día	Cada 8 días	Cada 5.5 días
Disponibilidad	Libre acceso	Libre acceso	Programa de Educación e Investigación de Planet Labs (2020)
Año de lanzamiento	1999	1999 y 2013	2009
Rango de datos	2013/10/03 - 2014/04/05 (184 días)		30/12/2013
Aplicación	Construcción de matrices espaciotemporales, generación de índices de vegetación y extracción de métricas fenológicas.		Segmentación de parcela y generación de índices de vegetación

3. Datos

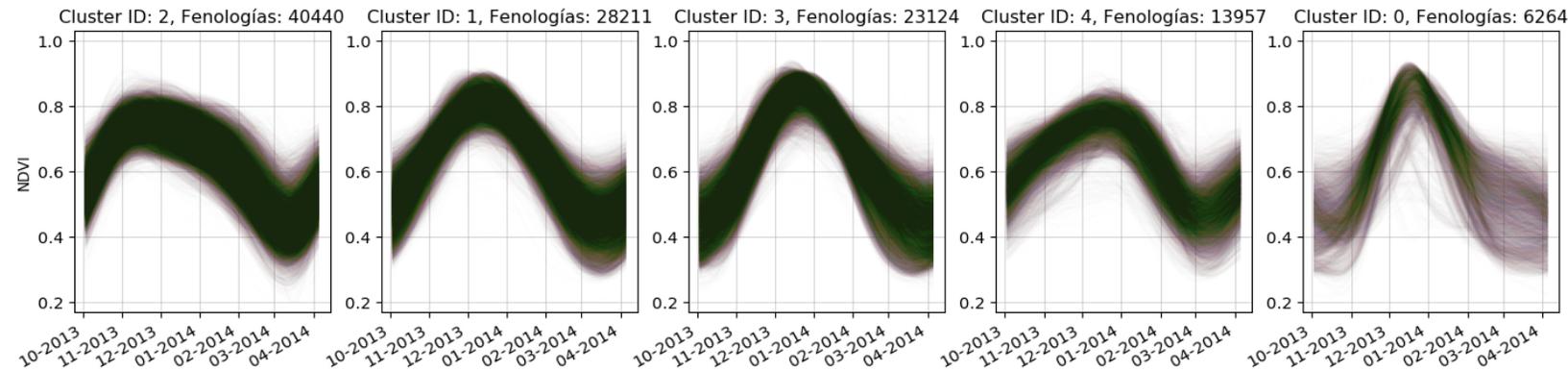
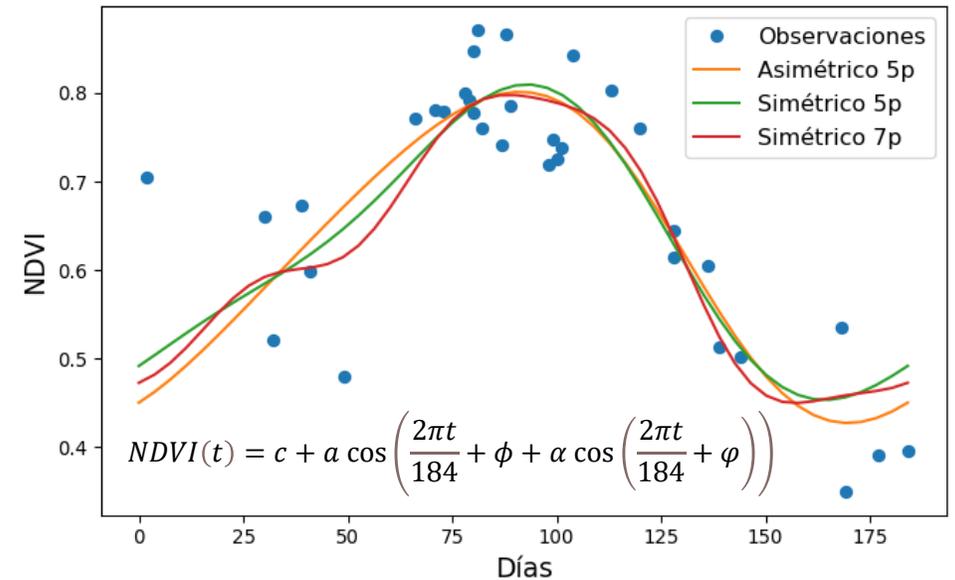
Procesamiento de imágenes satelitales

1. Preprocesamiento de las imágenes y construcción de series espaciotemporales

- Reproyección, evaluación de píxeles y corrección de imágenes
- Remuestreo de 250m a 30m (ej. fusión de imágenes)
- Cálculo del NDVI y construcción de series espaciotemporales (matrices tridimensional)

2. Ajuste de modelo armónico y clasificación de la superficie entre cultivos y pastos

- Modelo armónico asimétrico (Carrao et al., 2010)
- Fenología (cultivos transitorios)
- Clasificación entre cultivos y no cultivos con k-Shape (Paparrizos & Gravano, 2015)



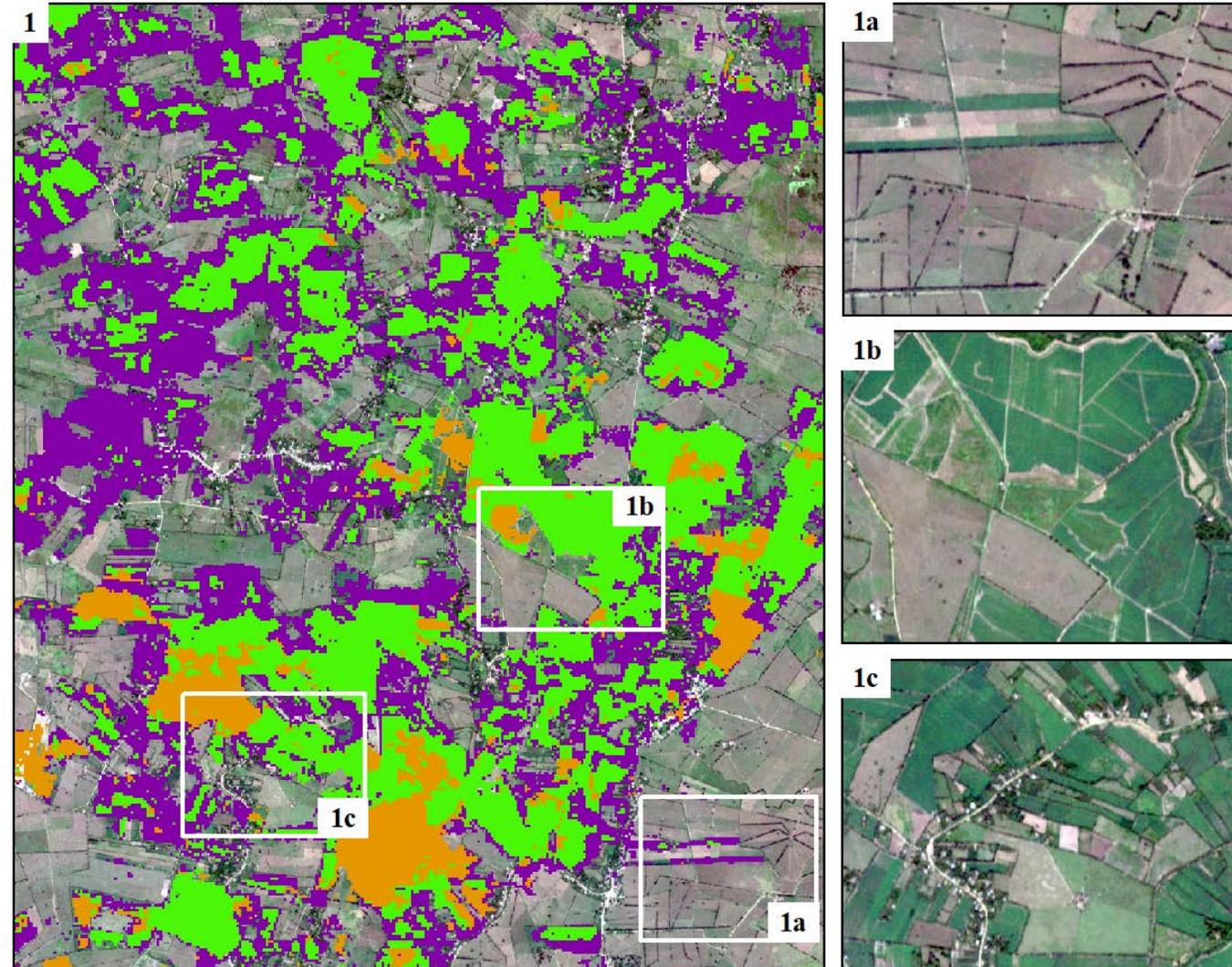
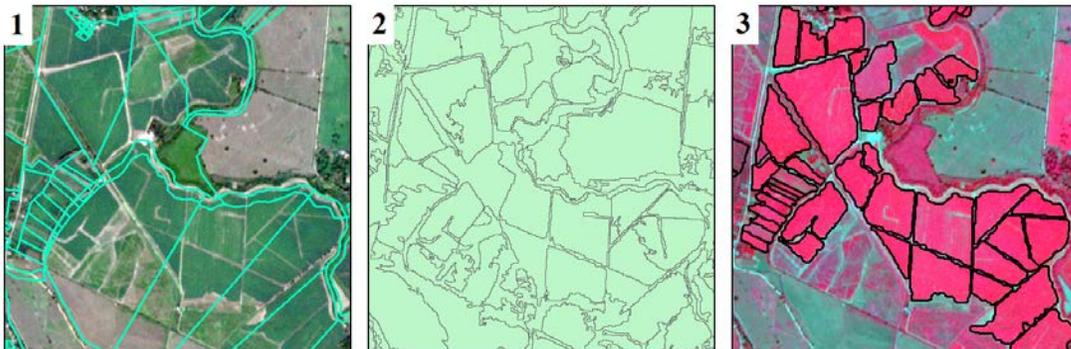
Total de píxeles 112.362 por imagen por 184 días (≈20 millones de observaciones)

3. Datos

- Problemas con los datos georreferenciados del CNA: sesgo en la distribución espacial, coordenadas repetidas, centroide de una UPA no de la parcela y polígonos de UPA confidenciales.
- Resultados validados por inspección visual

3. Segmentación de parcelas o lotes agrícolas

- Problemas con predio catastral rural (IGAC)
- Segmentación de imágenes con Mean-Shift (Michel & Grizonnet, 2015)



≈4 millones de píxeles por 5 bandas (≈20 millones de observaciones)

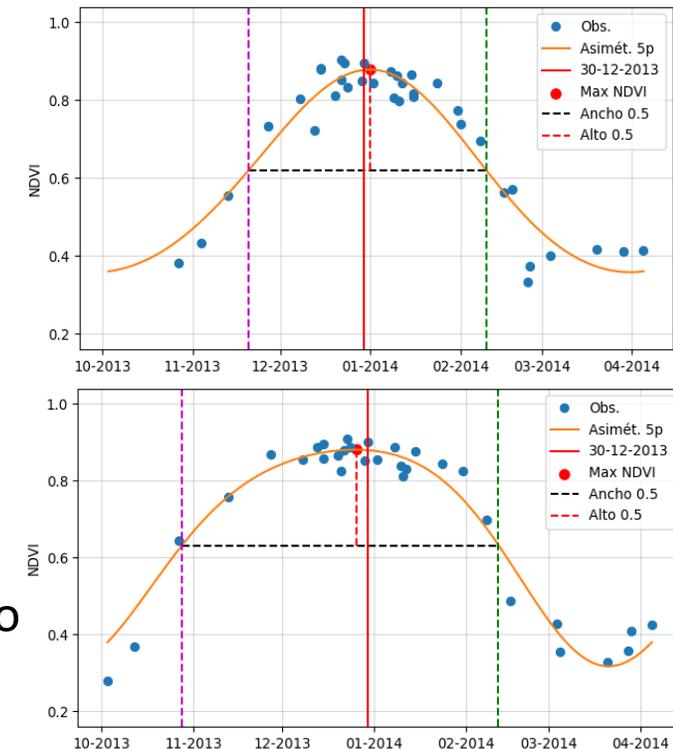
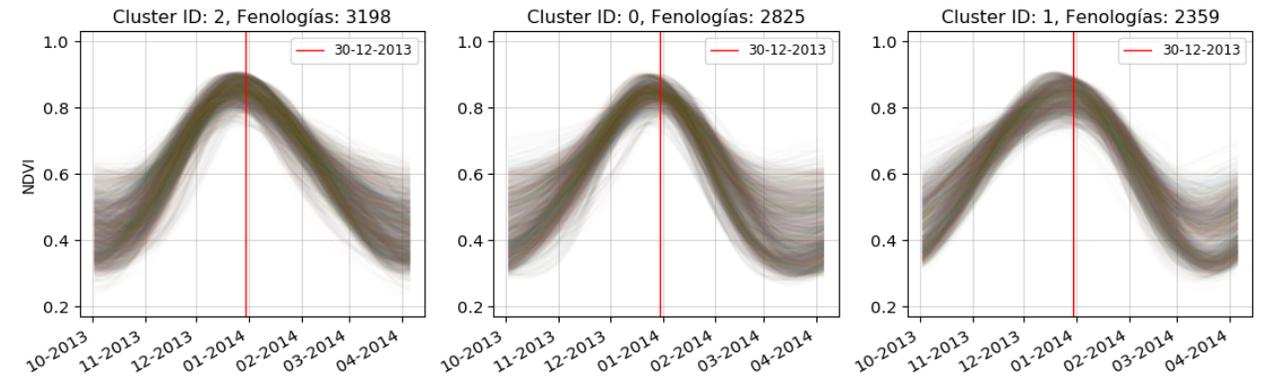
3. Datos

4. Filtro de predios con referencia al 30/12/2013 y clasificación de cultivos

- NDVI a 5m con imagen de RapidEye
- Parcelas cuyo pico de fenología estuviera a una diferencia absoluta de 15 días del 30/12/2013.
- Muestra final de parcelas de 292, con rango de píxeles por parcela de 1-188 a resolución de 30 m (900 m²) o de 19-8021 a 5 m (o 25 m²).
- k-Shape (de nuevo) para diferenciar entre cultivos (fenologías de algodón)

5. Proxy de rendimiento

- Las imágenes no pueden detectar el peso del algodón (solo proxies, ej. NDVI)
- Relación empírica entre el índice verde de área foliar (gLAI) y el NDVI
- Cálculo de rendimiento: encuestas, simulación, modelos estadísticos, etc.
- Sin datos de referencia, se considera NDVI como proxy de rendimiento, ej: máximo crecimiento alcanzado por la fracción del cultivo (pico de fenología)

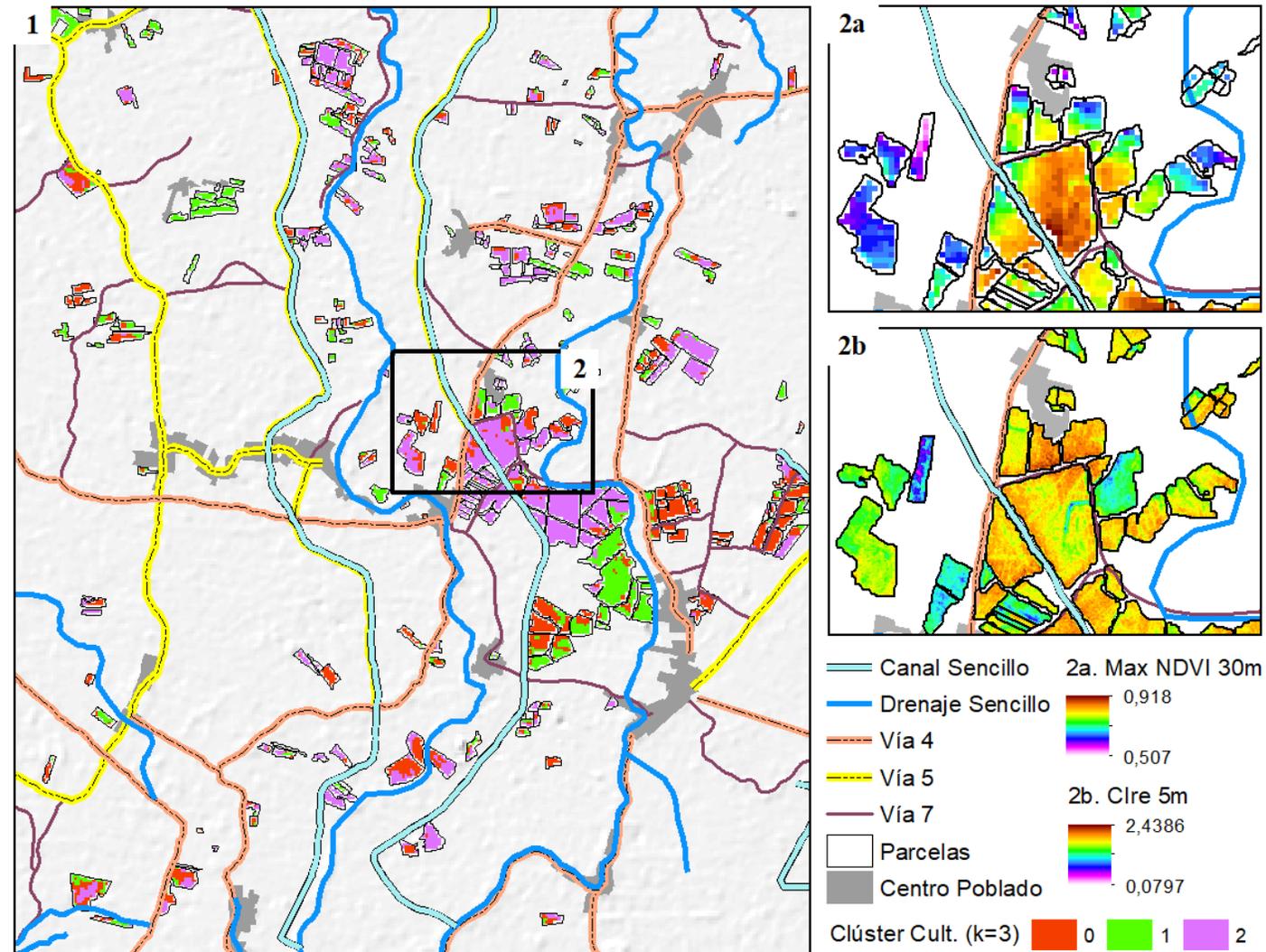


3. Datos

Otros datos

- Muestra final de parcelas (clúster de cultivos $k=3$): exposición directa a ríos que alteran la fertilidad del terreno; cercanía a canales que controlan excesos de agua causados por inundaciones; y altitud y pendiente del terreno que influyen sobre el nivel de erosión del suelo.
- Los ríos y los canales a escala de 1:100.000 del portal de Datos Abiertos del IGAC
- Altitud y la pendiente de un Modelo de Elevación Digital (DEM) a 30 m (AW3D30)
- Vías a escala de 1:100.000 del portal de Datos Abiertos del IGAC (ej. acceso a las parcelas)
- Caseríos o centros poblados del Geoportal del DANE.

NOTA: Se evitaron píxeles mixtos.



4. Estrategia empírica

Forma reducida de la relación tamaño-productividad utilizada por Bevis & Barrett (2020):

$$Y_{pft} = \frac{G_{pft}}{A_{pft}} = \alpha A_{pft} + \beta \frac{P_{pft}}{A_{pft}} + \lambda_{pf} + \lambda_t + \varepsilon_{pft} \quad (4)$$

Bevis & Barrett (2020)

Panel de datos parcela-hogar en dos periodos de tiempo (10 años) y área por GPS y calidad del suelo por análisis de laboratorio.

Esta investigación

Panel de datos píxel-parcela (fijo en el tiempo) para medir efecto de borde de forma directa. Controla por variables que no varían dentro de la parcela. Pero efectos fijos de parcela eliminan el área.

Modelo 5 a nivel de píxel:

$$Max_NDVI_{ijk} = \beta_1 \log(dist_borde_lote_{ijk}) + VT'_{ijk} \gamma + VA'_{ijk} \delta + \alpha_j + \tau_k + \varepsilon_{ijk} \quad (5)$$

Modelo 6 a nivel de píxel (sin incluir efectos fijos de parcela):

$$Max_NDVI_{ijk} = \beta_1 \log(area_lote_k) + \beta_2 \log(dist_borde_lote_{ijk}) + \alpha_j + \varepsilon_{ijk} \quad (6)$$

Modelo 7 a nivel de parcela:

$$prom_Max_NDVI_k = \beta_1 \log(area_lote_k) + \beta_2 \log(prom_dist_borde_lote_k) + \varepsilon_k \quad (7)$$

i, j y k son índices de píxel, clúster de formas y parcela, VT y VA son variables de terreno y accesibilidad, α_j y τ_k son efectos fijos de clúster y de parcela

5. Resultados

Tabla 2

Estimación del modelo 5 (variable dependiente *Max_NDVI*)

Variable	β Pooled	β FE	β 2FE	β 2FE	β 2FE
<i>log(dist_borde_lote)</i>	0,0198*** (0,00081)	0,00506*** (0,000767)	0,00498*** (0,00092)	0,00509*** (0,00094)	0,00488*** (0,00102)
VARIABLES DE TERRENO					
<i>elevacion</i>				-0,00034 (0,00029)	-0,00020 (0,00019)
<i>pendiente</i>				-0,00015 (0,00044)	-0,00011 (0,00043)
<i>log(dist_dren_senc)</i>				-0,00559*** (0,00152)	-0,00623*** (0,00138)
<i>log(dist_canal_senc)</i>				0,00696* (0,00350)	0,00677 (0,00366)
VARIABLES DE ACCESIBILIDAD					
<i>log(dist_cpob)</i>					0,00627 (0,00360)
<i>log(dist_via4)</i>					0,00533* (0,00266)
<i>log(dist_via5)</i>					0,00658 (0,00395)
<i>log(dist_via7)</i>					-0,00073 (0,00103)
No. Pixeles 30m (obs.)	4708	4708	4708	4708	4708
No. Lote (parcelas)	-	239	239	239	239
No. Clústeres de formas	-	-	3	3	3
Pixeles por lote	-	2 - 188	-	-	-
Pixeles por lote y clúster	-	-	2 - 160	2 - 160	2 - 160
R2	0,1023	0,0408	0,0409	0,0963	0,1446
Log-likelihood	9768,01	14760,42	14841,38	14981,58	15110,86
Tipo de errores estándar	<i>White</i>	<i>Agrupados en 1 dimensión</i>	<i>Agrupados en 2 dimensiones</i>	<i>Agrupados en 2 dimensiones</i>	<i>Agrupados en 2 dimensiones</i>
Variable de agrupamiento de errores robustos	-	Lote	Lote y Clust3	Lote y Clust3	Lote y Clust3

Nota: Errores estándar robustos en paréntesis. FE hace referencia a Efectos Fijos y FE2 a Efectos Fijos en dos niveles.

Lote es igual a parcela. * p<0.05, ** p<0.01, *** p<0.001

Tabla 3

Estimación del modelo 6 (variable dependiente *Max_NDVI*)

Variable	β Pooled	β Pooled
<i>log(area_lote)</i>	0,01728*** (0,00054)	0,01425*** (0,00064)
Efecto de borde		
<i>log(dist_borde_lote)</i>		0,00872*** (0,00092)
No. Pixeles 30m (obs.)	4708	4708
R2	0,1743	0,1887
Log-likelihood	9964,96	10006,78
Tipo de errores estándar	<i>White</i>	<i>White</i>

Nota: Errores estándar en paréntesis. * p<0.05, ** p<0.01, *** p<0.001

Tabla 4

Estimación del modelo 7 (variable dependiente *prom_Max_NDVI*)

Variable	β	β	β
<i>log(area_lote)</i>	0,01361*** (0,00250)	-0,00677 (0,00710)	-0,00208 (0,00482)
<i>log(ratio_perimetro_area)</i>		-0,0421** (0,01376)	
<i>log(prom_distancias_borde)</i>			0,04664*** (0,01237)
<i>Intercepto</i>	0,8323*** (0,00269)	0,80218*** (0,01022)	1,0079*** (0,04664)
Lotes (parcelas)	246	246	246
R2 ajustado	0,1046	0,1343	0,1506
F-statistic	29,63***	20***	22,72***

Nota: Errores estándar en paréntesis. * p<0.05, ** p<0.01, *** p<0.001. Ratio es igual a cociente.

6. Robustez

El modelo 5 tienen por lo menos **3 inconvenientes**:

- El NDVI se satura asintóticamente con valores del LAI superiores a 3 (Xie et al., 2018)
- El píxel de 30m es muy grande en relación con el tamaño de las parcelas de la región de estudio
- Las vías, ríos, canales, etc. están sujetos a errores de medición dado que fueron delimitados a una escala de análisis mucho más grande a la requerida



El modelo 7 tienen por lo menos **1 inconveniente**:

- Las parcelas no están asociadas a las fincas o UPA a las que pertenecen. No es posible controlar por efectos fijos de finca u hogar (ej. por fallas en los mercados)



Por lo anterior y el manejo de variables proxy es prematuro afirmar un **efecto causal** entre las variables que describen el efecto de borde y menos entre las que explican la relación tamaño-productividad de las parcelas de esta región.

7. Conclusiones

- Se probó la **existencia del efecto de borde** para una muestra de parcelas dedicadas al cultivo de algodón y ubicadas en el departamento de Córdoba (Colombia)
- Se encontró que el crecimiento de los doseles de las plantas (proxy de rendimiento) **era menor** hacia los bordes de las parcelas (contrario a Bevis y Barret, 2020)
- No se pudo establecer la **covariable** que explica este efecto de borde
- Se identificó que, a nivel de píxel o a nivel de parcela, la relación entre el tamaño-productividad (biofísica) era **directa** en vez de inversa
- Similar a Bevis & Barrett (2020), esta relación se explicaba por el **cociente perímetro-área** o por el **promedio de distancias al borde de la parcela** (efecto de borde indirecto)
- **Política:** aún es **prematureo emitir recomendaciones** agrícolas para evitar el bajo crecimiento de las plantas hacia el borde de las parcelas o de política económica que fomente una redistribución o no de la tierra
- Debido a los problemas de las diferentes fuentes de información disponibles, se necesitan protocolos que faciliten la **integración** entre encuestas de campo y fuentes alternativas de información (ej. catastro multipropósito)
- **Investigaciones futuras:** expansión temporal de la base de datos, y verificación de los cultivos y refinamiento de las proxies de rendimiento con datos de campo.

¡Gracias por su atención!

8. Bibliografía

- Assuncao, J. J., & Braido, L. H. B. (2007). Testing Household-Specific Explanations for the Inverse Productivity Relationship. *American Journal of Agricultural Economics*, 89(4), 11.
- Barrett, C. B. (1996). On price risk and the inverse farm size-productivity relationship. *Journal of Development Economics*, 51(2), 193–215.
- Barrett, C. B., Bellemare, M. F., & Hou, J. Y. (2010). Reconsidering Conventional Explanations of the Inverse Productivity-Size Relationship. *World Development*, 38(1), 88–97.
- Benjamin, D. (1995). Can unobserved land quality explain the inverse productivity relationship? *Journal of Development Economics*, 46(1), 51–84.
- Bevis, L. E., & Barrett, C. B. (2020). Close to the edge: High productivity at plot peripheries and the inverse size-productivity relationship. *Journal of Development Economics*, 143, 102377.
- Carletto, C., Jolliffe, D., & Banerjee, R. (2015). From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. *Journal of Development Studies*, 51(2), 133–148.
- Carletto, C., Savastano, S., & Zezza, A. (2013). Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. *Journal of Development Economics*, 103(1), 254–261.
- Carrao, H., Gonçalves, P., & Caetano, M. (2010). A Nonlinear Harmonic Model for Fitting Satellite Image Time Series: Analysis and Prediction of Land Cover Dynamics. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4), 12.
- Donaldson, D., & Storeygard, A. (2016). The View from Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives*, 30(4), 171–198.
- FAO. (2017). *Productivity and Efficiency Measurement in Agriculture: Literature Review and Gaps Analysis*.
- Feder, G. (1985). The relation between farm size and farm productivity: The role of family labor, supervision and credit constraints. *Journal of Development Economics*, 18(2–3), 297–313.
- Gáfaró, M., Ibañez, A. M., & Zarruk, D. (2012). Equidad y eficiencia rural en Colombia: una discusión de políticas para el acceso a la tierra. *Documentos CEDE*.

8. Bibliografía

- Helfand, S. M., & Levine, E. S. (2004). Farm size and the determinants of productive efficiency in the Brazilian Center-West. *Agricultural Economics*, 31(2–3), 241–249.
- Lamb, R. L. (2003). Inverse productivity: land quality, labor markets, and measurement error. *Journal of Development Economics*, 71(1), 71–95.
- Maus, V., Câmara, G., Appel, M., & Pebesma, E. (2017). dtwSat: Time-Weighted Dynamic Time Warping for Satellite Image Time Series Analysis in R. *Journal of Statistical Software*, 88, 1–30.
- Michel, J., & Grizonnet, M. (2015). State of the Orfeo Toolbox. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 1336–1339). IEEE.
- Murcia, C. (1995). Edge effects in fragmented forests: implications for conservation. *Trends in Ecology & Evolution*, 10(2), 58–62.
- Ospina Sierra, S. (2016). *Eficiencia y tamaño del predio agrícola en Colombia*. Universidad de los Andes. Retrieved from <http://hdl.handle.net/1992/13746>
- Paparrizos, J., & Gravano, L. (2015). K-Shape: Efficient and Accurate Clustering of Time Series. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data* (pp. 1855–1870). New York, NY, USA: Association for Computing Machinery.
- Planet Labs. (2020). Education and research program. Retrieved April 23, 2020, from <https://www.planet.com/markets/education-and-research/>
- Raatz, L., Bacchi, N., Pirhofer Walzl, K., Glemnitz, M., Müller, M. E. H., Joshi, J., & Scherber, C. (2019). How much do we really lose?—Yield losses in the proximity of natural landscape elements in agricultural landscapes. *Ecology and Evolution*, 9(13), 7838–7848.
- Sen, A. K. (1966). Peasants and Dualism with or without Surplus Labor. *Journal of Political Economy*, 74(5), 425–450.