

Expl Agric.: page 1 of 23 © Cambridge University Press 2016. Este es un artículo de acceso abierto, distribuido bajo los términos de la licencia Creative Commons Attribution (<http://creativecommons.org/licenses/by/4.0/>), que permite su irrestricta reutilización, distribución y reproducción a través de cualquier medio, siempre y cuando se cite debidamente la obra original.

doi:[10.1017/S0014479716000193](https://doi.org/10.1017/S0014479716000193)

MÁS ALLÁ DE LOS PROMEDIOS: NUEVOS ENFOQUES PARA ENTENDER LA HETEROGENEIDAD Y EL RIESGO DEL ÉXITO O FRACASO DE LA TECNOLOGÍA EN LOS SISTEMAS AGRÍCOLAS DE PEQUEÑOS PRODUCTORES

Por BERNARD VANLAUWE^{†‡}, RIC COE[§] y KEN E. GILLER[¶]

†Área de Investigación de Gestión de Recursos Naturales, Instituto Internacional de Agricultura Tropical (CIAT), Apartado Postal 30772, Nairobi 00100, Kenia, §Centro Agroforestal Mundial (ICRAF), Nairobi, Kenia y Centro de Servicios Estadísticos de la Universidad de Reading, Reading, Reino Unido y ¶Sistemas de Producción Vegetal, Universidad de Wageningen (WUR), Apartado Postal 430, 6700 AK Wageningen, Países Bajos

(Aceptado el 10 de febrero de 2016)¹

RESUMEN

En los últimos años, muchos estudios han demostrado la heterogeneidad del entorno de producción de los pequeños agricultores. Sin embargo, la investigación agronómica para el desarrollo (R4D), cuyo objetivo es identificar y probar opciones para aumentar la productividad, no ha adaptado consistentemente sus enfoques a estas condiciones de heterogeneidad. En este artículo se describen los desafíos que enfrenta la investigación, destacando la importancia de la variación al evaluar la implementación de las recomendaciones para el manejo del suelo, integrando dentro de la formulación de recomendaciones los aspectos de la gestión de riesgos en la producción y proponiendo enfoques alternativos para implementar la investigación agronómica para el desarrollo (R4D). Para ilustrar los enfoques se utilizan dos pruebas pareadas multilocalizadas realizadas en finca. Cada una de ellas tiene un tratamiento de control sin insumos y un tratamiento con aplicación de fertilizante para el maíz en el oeste de Kenia y para el frijol en el este de Ruanda. La diversidad de respuestas al tratamiento debería acogerse, en lugar de evitarse, para así entender mejor el contexto actual y su relación con la gestión anterior.

INTRODUCCIÓN

La investigación en agronomía se basa en una tradición establecida desde hace mucho tiempo. Ha traído aumentos significativos en los rendimientos de todos los principales cultivos alimenticios y

[‡] Autor correspondiente. IITA Kenia, c/o ICIPE, Off Thika Highway, PO Box 30772, Nairobi 00100, Kenia, Email:b.vanlauwe@cgiar.org

¹ Traducido de inglés a español por Cristina Jaramillo Lopera y Adriana López Henao

otros productos básicos, por ejemplo multiplicando por diez la productividad de los principales cereales, al pasar de 1 t ha⁻¹ a comienzos del siglo XIX a más de 10 t ha⁻¹ actualmente en algunas fincas. Las primeras pruebas experimentales establecidas por Lawes y Gilbert en Rothamsted en la década de 1840, varios de los cuales aún continúan hoy, consistían en una sola parcela de cada tratamiento (Poulton, 2006). No fue sino hasta la década de 1920 que los métodos investigativos para la experimentación vivieron una revolución gracias al trabajo de Ronald Fisher. Fisher introdujo por primera vez el *Análisis de la varianza* (ANOVA) en 1925 en su libro *Métodos estadísticos para investigadores* (Fisher, 1925). Los principios de diseño que Fisher introdujo en su libro *El Diseño de experimentos* proporcionaron métodos para garantizar la eficiencia, eliminar el sesgo en la comparación de tratamientos y brindar una medida de precisión de los resultados incluyendo el bloqueo para minimizar los efectos de la variación no explicada (Fisher, 1935). Rápidamente el ANOVA de Fisher y su comparación de las medias se volvieron el método estándar para el análisis de los experimentos agronómicos diseñados para examinar los efectos de factores únicos o múltiples y sus interacciones. Los métodos de diseño y análisis están basados en el supuesto de que existe un único efecto “verdadero” de tratamiento a estimar. Tales diseños tenían como objetivo minimizar la variación no explicada, o la variación no explicada por los tratamientos aplicados. Normalmente los experimentos se llevaban a cabo en una sola o en muy pocas locaciones, la mayoría bajo condiciones bien manejadas, “en la estación”.

Los métodos de Fisher se convirtieron en la base para la experimentación agrícola y sin duda contribuyeron enormemente a los programas de investigación que condujeron a la Revolución Verde en América Latina y Asia. Durante la época de la colonia en África, las estaciones de investigación agrícola habían sido ubicadas en los mejores suelos, principalmente para ofrecer buenas locaciones para el cultivo de productos básicos (*commodities*) específicos destinados a la siembra y la explotación agrícola a gran escala. Por ejemplo, en Tanzania la estación de investigación de café estaba ubicada en Lyamungu, en las laderas del Kilimanjaro, la estación de investigación del algodón en Ukiriguru, al sur del lago Victoria, y la estación de investigación de horticultura en Tengeru, Arusha. Estas estaciones de investigación sirvieron a su propósito cuando el asunto era la propagación a gran escala de estos productos básicos de altos insumos, brindando condiciones cercanas a las ideales para los productos básicos específicos para los cuales fueron establecidos. Unas cuantas estaciones de investigación tenían una ubicación que no era la ideal, por ejemplo Peter Le Mare descubrió que la estación de Namulonge en Uganda estaba situada arriba de un filón de rocas que contenían manganeso, con propiedades del suelo peculiares en términos de disponibilidad de fósforo (Le Mare, 1968). El enfoque se centró en gran medida en el mejoramiento genético (por ejemplo de rendimiento, resistencia a las enfermedades, variedades enanas que pudieran responder a N, o estrés por sequía). La experimentación agronómica desempeñó un papel fundamental de apoyo para optimizar el rendimiento de los nuevos genotipos de cereales enanos, al intensificar la densidad vegetal, las tasas de fertilización, etc. Los diseños estadísticos basados en el enfoque original de Fischer, con muchas extensiones y adiciones elegantes a estos métodos (por ejemplo Cochran y Cox, 1957) demostraron ser herramientas poderosas para clarificar los roles de factores específicos o de sus interacciones. El modelo de desarrollo agrícola fue de investigación y demostración. La experimentación era el

dominio de los investigadores y la demostración le pertenecía al departamento de extensión, cuyo rol se limitaba a difundir los resultados de la investigación experimental para convencer sobre su adopción a un amplio grupo de productores (Maat y Glover, 2012). Algunas investigaciones agronómicas iniciales reconocieron la necesidad de orientar específicamente las recomendaciones a diferentes zonas agroecológicas, tales como las investigaciones pioneras para diferenciar las curvas de respuesta a los fertilizantes (Scaife, 1968). En Kenia, el Proyecto de Recomendaciones de Uso de Fertilizantes (FURP por sus iniciales en inglés, 1994) intentó desagregar las respuestas a los fertilizantes pero lo hizo por un distrito administrativo en lugar de basar las recomendaciones en factores que posiblemente interactuaran con la respuesta a los fertilizantes. Otros reconocieron desde el inicio la necesidad de investigación en finca, para identificar los factores que limitaban la productividad de las fincas de los pequeños agricultores, tales como las deficiencias de K (Anderson, 1974).

El pobre desempeño de las tecnologías en el momento en que se encontraban fuera de las estaciones experimentales, y su escasa adopción en las fincas de los pequeños productores, condujeron al advenimiento de la investigación sobre sistemas agrícolas (Collinson, 2000) y la selección participativa de variedades (Sperling and Ashby, 2000). Estos enfoques buscaban entender el contexto dentro del cual se desarrolla la producción de cultivos e involucrar a los productores desde los inicios del proceso de investigación. La necesidad de esto era doble: en primer lugar, a menudo las variedades u opciones para el manejo agronómico no coincidían con las preferencias o intereses de los agricultores; y segundo, con frecuencia las opciones mostraban un menor rendimiento en las tierras de los agricultores que el que tenían bajo las condiciones de las estaciones de investigación, en las cuales estaban monitoreadas de cerca. Las principales razones para el fracaso de las tecnologías fueron la inherente escasa fertilidad del suelo en las fincas de los pequeños productores (Waddington y col., 1998) y la excesiva demanda de mano de obra y otras restricciones de tipo social de muchas de las tecnologías propuestas. El caso de cultivos en callejones fue documentado por Carter (1995).

Kofi Anan resaltó la necesidad de incrementar la productividad agrícola en África mediante una “Revolución Verde singularmente Africana” que reconozca la diversidad de gentes, suelos y prácticas agrícolas (Tittonell y col., 2011). La situación en África difiere mucho de las condiciones que apoyaron la Revolución Verde en Asia, en términos de diversidad de las condiciones agroecológicas, falta de oportunidades para el riego generalizado, y apoyo de los gobiernos en cuanto a políticas (Djurveldt y col., 2005). Por lo tanto, la investigación agronómica también debe hacer frente a la diversidad de condiciones en las cuales se desarrolla la agricultura de pequeños productores (Buerkert y col., 2001), de la misma manera que los servicios de asesoramiento deben ofrecer las opciones más idóneas para adaptarse a las capacidades locales y a un entorno normativo más amplio (Birner y col., 2009).

Podría argumentarse que la mayoría de los aspectos básicos de la agronomía y del manejo del cultivo están bien establecidos. A medida que el foco de la investigación se ha desplazado cada vez más hacia un entendimiento de la productividad de las variedades de cultivos o el manejo

agronómico bajo las condiciones de los productores, las preguntas de investigación han evolucionado. En esencia, existe cada vez más conciencia de que las preguntas a las que se ve enfrentada la investigación agronómica también han cambiado. Más allá de entender los impactos de factores individuales, necesitamos entender las interacciones entre una amplia gama de factores que interactúan para determinar la productividad del cultivo. Esto requiere pruebas generalizadas de las tecnologías, ya que el sesgo en la selección de los sitios y un manejo “estandarizado” tienden a ignorar o a controlar los factores que determinan la respuesta del cultivo. Incluso los diseños incompletos solo pueden incluir un pequeño número de factores antes de que los experimentos se vuelvan inmanejables si se llevan a cabo en muchos sitios. Esto ha llevado a que la distinción entre los experimentos y las demostraciones sea cada vez más difusa. Mientras que la investigación agrícola ha tendido a recomendaciones muy generales que puedan ser aplicadas rápidamente en grandes áreas, cada vez se reconoce más el hecho de que la heterogeneidad de agroecologías y de fincas y productores requiere soluciones adaptadas y matizadas localmente, enfoques idóneos hechos a la medida (Giller y *col.*, 2011). La investigación sobre el manejo integrado de la fertilidad del suelo (MIFS) ha acogido esta visión enfocándose en la adaptación local de tecnologías para mejorar la fertilidad del suelo (Vanlauwe y *col.*, 2015). Antes de llegar a las “recomendaciones” es necesario entender la frecuencia de respuesta y la frecuencia de la magnitud de la respuesta a las diversas intervenciones. Se han dedicado relativamente pocos esfuerzos a la tarea de formular hipótesis o preguntas precisas con respecto a los efectos de tratamientos agronómicos en fincas, o a desarrollar métodos para analizar los efectos de tratamientos experimentales sobre un gran conjunto de datos. Llevando esto a una etapa más avanzada, es esencial comprender los principales factores que rigen la respuesta a un determinado tratamiento, con el fin de brindar información a los agricultores sobre la probabilidad de obtener beneficios sustanciales o económicos si invierten en determinada tecnología.

En este artículo: (i) describimos los desafíos que supone dirigirse a entornos de pequeños productores heterogéneos y los diversos sistemas agrícolas de pequeños productores; (ii) resaltamos la importancia de la variación cuando se evalúa el desempeño de las recomendaciones sobre manejo del suelo; (iii) integramos aspectos de gestión de riesgos en la producción con la formulación de recomendaciones y (iv) describimos las consecuencias de lo anterior para la implementación de la investigación agronómica a gran escala.

LAS REALIDADES DE LOS SISTEMAS AGRÍCOLAS DE PEQUEÑOS PRODUCTORES EN EL ÁFRICA SUBSAHARIANA

Los diseños de investigación agronómica “tradicionales” estaban enfocados en controlar y minimizar la variación ocasionada por factores no incluidos en la estructura del tratamiento. Cuando se evalúa el impacto de los tratamientos realizados bajo la administración del agricultor, los factores en la finca interactúan con los tratamientos prioritarios, bien sea agronómicamente (por ejemplo el momento e intensidad de prácticas agronómicas como la siembra y el deshierbe) o

económicamente (por ejemplo los precios de los insumos y los productos para el producto básico cultivado) de manera que dichos factores pueden fácilmente enmascarar o reducir el desempeño potencial de estos tratamientos. Los entornos de sistemas agrícolas de pequeños productores en el África subsahariana se caracterizan por (i) condiciones de fertilidad del suelo que varían dentro de distancias cortas, aún dentro de la misma finca, a lo que nos referimos a menudo como gradientes de fertilidad del suelo (Buerkert y *col.*, 2001; Tittonell y *col.*, 2005; Zingore y *col.*, 2007), (ii) variaciones en el acceso a los recursos por parte de las familias campesinas dentro de las mismas comunidades (Giller y *col.*, 2011; Tittonell y *col.*, 2010) y (iii) condiciones variables de habilitación para la intensificación agrícola, incluyendo el acceso a insumos agrícolas, mercados y servicios de extensión (Sumberg, 2005).

Dentro de las fincas, la interacción de los procesos de formación del suelo y el manejo preferencial han dado como resultado tierras con diferentes condiciones de fertilidad del suelo. Algunos campos están tan degradados que la aplicación de fertilizantes estándar no se traduce en incrementos sustanciales en el crecimiento del cultivo; son conocidos a menudo como suelos no reactivos. La aplicación de opciones de manejo consideradas como “las óptimas” dentro de condiciones de fertilidad del suelo tan heterogéneas comúnmente resulta en un amplio rango de respuestas, desde prácticamente ningún efecto hasta incrementos que se aproximan al rendimiento potencial para un entorno en particular (Vanlauwe y *col.*, 2006). Nosotros definimos las alternativas “óptimas” como aquellas que, con la comprensión actual de los objetivos de los pequeños productores y el entorno de producción, se prevé que tengan buena probabilidad de alcanzar esos objetivos. El componente “objetivos de los pequeños productores” en esta definición involucra aspectos de cantidad y magnitud de riesgos, resultados a corto plazo versus largo plazo, y múltiples opciones (por ejemplo diferentes opciones para diferentes tipos de agricultores).

Dentro de las comunidades, los agricultores tienen (i) diferentes posibilidades de acceso a los recursos de producción como tierra, mano de obra y dinero en efectivo, (ii) diferentes objetivos de producción, entre los que se encuentran alimentos para la subsistencia y productos para el mercado, (iii) diferentes capacidades para absorber los riesgos inherentes a las prácticas de manejo alternativas, siendo los hogares más pobres los que más tratan de evitar los riesgos y (iv) diferentes actitudes hacia la explotación agrícola y al rol que esta desempeña dentro de todos los medios de subsistencia. Estas condiciones a nivel de fincas afectan las elecciones de prácticas de manejo agronómico y los efectos que se esperan de los tratamientos considerados los más idóneos. Es ampliamente reconocido que el rendimiento de un cultivo es el resultado de las interacciones entre el genotipo del cultivo (G) el entorno dentro del cual existe el cultivo (E) y el manejo agronómico (M), lo cual plantea el marco de trabajo $G \times E \times M$. Dentro de entornos agrícolas a nivel subnacional, nacional o regional, las condiciones habilitadoras varían considerablemente, incluyendo el acceso a los mercados de insumos y productos, la infraestructura rural y las condiciones normativas. A su vez, las instituciones afectan la rentabilidad de la explotación agrícola y el interés potencial de las comunidades agrícolas en nuevas tecnologías. En este contexto, las instituciones se refieren a los factores sociales,

económicos, políticos y de infraestructura, entre otros, que influyen en el desempeño de tecnologías específicas. Ojiem y col. (2006) se referían a las interacciones entre estos factores como *el nicho socioecológico* de las tecnologías. Nelson y Coe (2014) lo describen en forma más general como la *interacción de opciones por contexto*.

Tras lo anterior y reconociendo que hay una multitud de dimensiones agrícolas que interactúan con los tratamientos “óptimos”, está claro que los enfoques tradicionales ofrecen un indicativo de los efectos potenciales de tratamiento en el mejor de los casos, y que los diseños multifactoriales que abarcan todos los posibles factores de influencia se vuelven prácticamente imposibles de diseñar y establecer. Esto reclama nuevos enfoques para implementar las pruebas agronómicas y analizar los datos obtenidos de dichas pruebas.

ESTUDIOS DE CASO

Para ilustrar los nuevos enfoques y herramientas analíticas utilizamos dos estudios de caso basados en pruebas en fincas en múltiples locaciones, realizadas con maíz en el oeste de Kenia y con fríjol en el este de Ruanda. La [Tabla 1](#) describe las áreas objetivo específicas, las condiciones ambientales y los detalles experimentales de las pruebas.

Utilizando enfoques estadísticos estándar, los resultados del estudio en Kenia mostraron un incremento en los rendimientos promedio del maíz después de la aplicación de fertilizante NPKS, pasando de 2.0 t ha⁻¹ a 3.6 t ha⁻¹ ([Figura 1a](#)), un incremento medio de 1.6 t ha⁻¹, con un error estándar de esta diferencia de medias de 0.3 t ha⁻¹. Normalmente esta diferencia sería reportada como “altamente significativa, $p < 0.001$ ”. De manera similar, el estudio en Ruanda mostró un incremento considerable en los rendimientos del fríjol tras la aplicación de fertilizante DAP, pasando de 0.7 t ha⁻¹ en el control a 1.1 t ha⁻¹ en el tratamiento con aplicación de fertilizante ([Figura 1b](#)), una diferencia de 0.4 t ha⁻¹ con un error estándar de la diferencia de medias de 0.1 t ha⁻¹ ($p < 0.001$). Los diagramas de caja-bigote indican los rangos de datos de rendimiento que se observaron ([Figura 1c](#) y [Figura 1d](#)).

Los efectos de tratamiento “significativos” que se describen en esta sección mediante la forma estándar de reportar los resultados, ocultan mucha información. Cada finca tiene dos parcelas con una repetición de cada tratamiento asignada aleatoriamente a cada parcela. El análisis habitual considera las fincas como bloques y utiliza un análisis de varianza (ANOVA) de dos vías para estudiar la importancia relativa de los efectos de tratamiento y calcular el efecto medio y su error estándar. Esto corresponde al uso del modelo

$$y_{ij} = m + f_i + t_j + \varepsilon_{ij}, \quad (1)$$

siendo y_{ij} la respuesta (por ejemplo el rendimiento) del tratamiento j en la finca i , m el rendimiento medio global, f_i son los efectos en la finca, t_j son los coeficientes de tratamiento, y ε_{ij}

los términos residuales, asumiendo que tiene una varianza constante de σ_2 . El interés se centra en el efecto medio del tratamiento

$$t = t_2 - t_1, \quad (2)$$

siendo t la diferencia entre los valores esperados de los dos tratamientos, t_2 el valor esperado del efecto del tratamiento 2, y t_1 el del tratamiento 1. Se asume que t es constante.

Tabla 1. Detalles seleccionados de los estudios de caso, incluyendo la estructura y alcance geográfico del tratamiento, y las condiciones agroecológicas. Ambos estudios de caso tuvieron un diseño multilocacional de dos tratamientos (control y tratamiento), sin que el control recibiera ningún insumo orgánico o mineral.

Estudio de caso (estación)	Alcance geográfico	Condiciones agroecológicas	Tipos de suelo	Prácticas comunes para los dos tratamientos	Insumos utilizados en las parcelas tratadas
Maíz en Kenia (1 ^{ra} temporada de cultivo de 2004)	Los distritos de Kenia: Vihiga (sub-locaciones de Ebusiloli y Emusutsi) y Siaya (sub-locaciones de Nyalugunga y Nyabeda)	Msnm: 1.200 – 1.600 m; precipitación anual: 1.400 – 1.800 mm (bimodal)	Nitisoles, Ferralsoles, Acrisoles	Variedad de maíz Híbrido HB513; plantada en 75 por 25 cm con 2 semillas por montículo, reducido después de 3 semanas	N, P, K, y S, aplicados a 80, 60, 60, y 24 kg de nutriente ha ⁻¹ , usando una combinación de urea (aplicada partida), super fosfato triple (SFT), super fosfato simple (SFS) y muriato de potasio (MOP), atado cerca de la línea de siembra e incorporado
Frijol en Ruanda (1 ^{ra} temporada de cultivo de 2009)	Distrito de Umutara (locaciones de Nyakigando Kabarore, Rugarama, Murambi); Distrito de Kibungo (locaciones de Gatore, Kabare), Distrito de Bugesera (locaciones de Musenyi, Mayange)	Msnm: 1.300 – 1.600 m; precipitación anual: 800 – 1100 mm (bimodal)	Ferralsoles húmicos, Acrisoles háplicos, Regosoles dístricos	Variedades de frijol: Bugesera: RWR1180, Kibungo: RWR1668 (Gatore) o Coltan (Kabare), Umutara: BRB194; plantado en 40 por 10 cm	N y P, aplicados a 36 y 46 kg de nutriente ha ⁻¹ como fertilizante de fosfato diamónico (DAP); aplicado al momento de sembrar en un surco paralelo a la línea de siembra, e incorporado

ANÁLISIS ALTERNATIVO DE LOS MISMOS DATOS

También se puede examinar el mismo conjunto de datos empleando métodos alternativos y revelando información adicional incorporada en ellos.

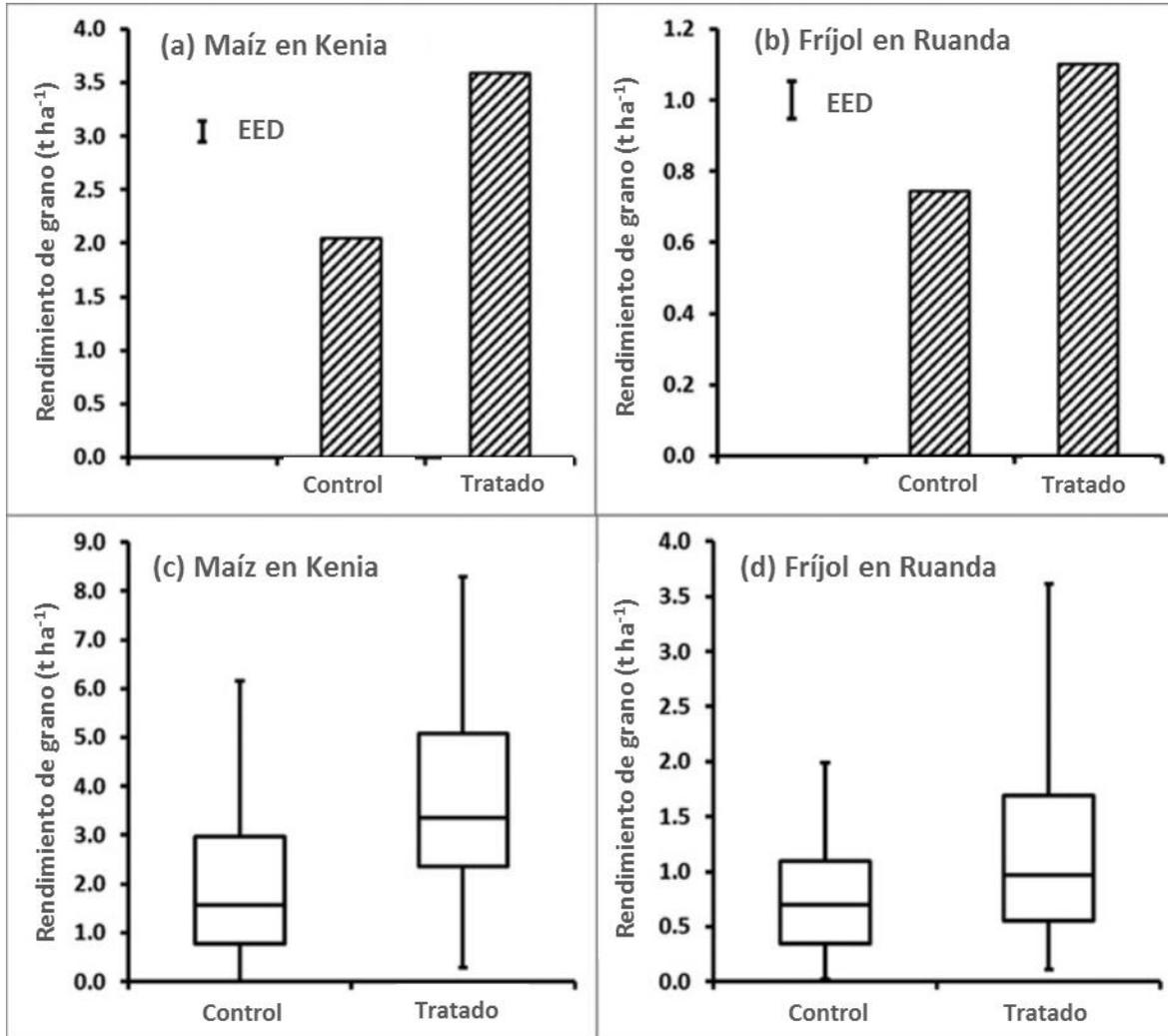


Figura 1. Rendimientos promedio (1a, 1b) y rendimientos presentados en formato de caja-bigote (1c, 1d) para los datos sobre el maíz de Kenia (1a, 1c) y el frijol de Ruanda (1b, 1d). 'EED' en 1a y 1b se refiere a 'Error Estándar de la Diferencia'.

Curvas de frecuencia acumulada y límites de confianza

El primer método se enfoca en la variabilidad de los efectos de tratamiento a lo largo de las fincas. El modelo (1) se reemplaza por

$$y_{ij} = m + f_i + t'_{ij} + \varepsilon_{ij}. \quad (3)$$

Aquí t'_{ij} es el parámetro para el tratamiento j en la finca i . El efecto de tratamiento en la finca i es

$$t'_i = t'_{i2} - t'_{i1}. \quad (4)$$

El mejor estimado que tenemos de estos efectos de tratamiento a nivel de la finca es la diferencia de rendimientos en cada finca

$$d_i = y_{i2} - y_{i1} = t'_i + \varepsilon'_i, \quad (5)$$

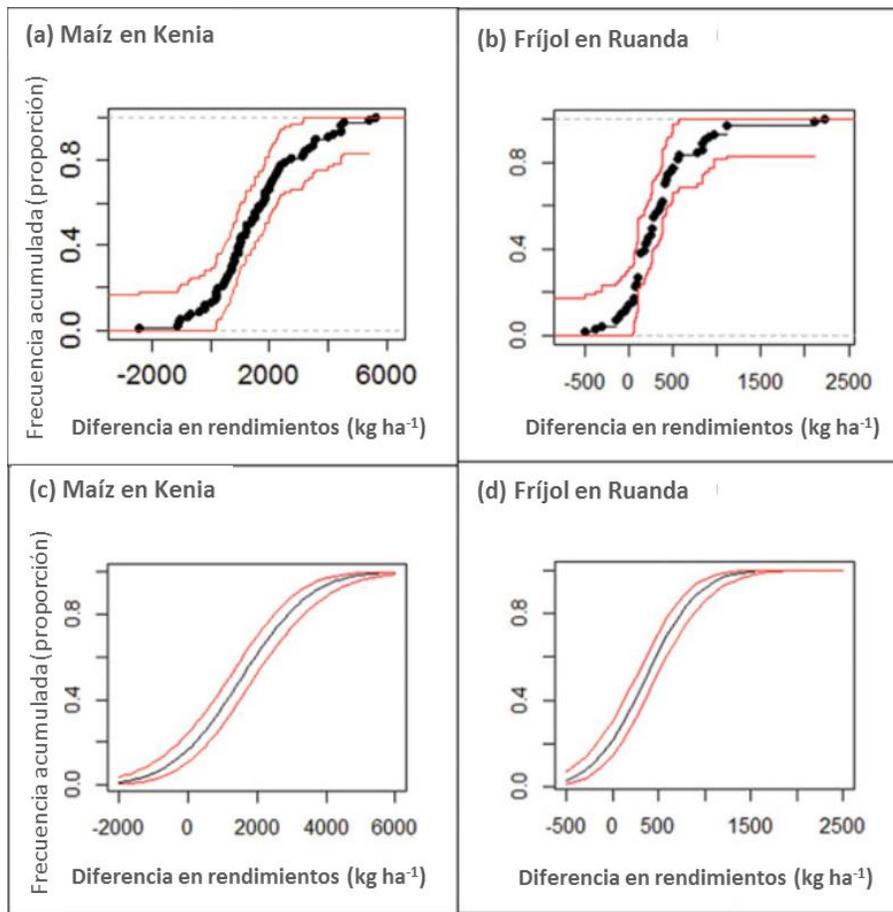


Figura 2. Intervalos de confianza dibujados alrededor de las curvas de frecuencia acumulada utilizando el estadístico D Kolmogorov –Smirnov (2a, 2b) y el modelo de distribución normal (2c, 2d) para el estudio del maíz en Kenia (2a, 2c) y el estudio del fríjol en Ruanda (2b, 2d).

donde

$$\varepsilon'_i = \varepsilon_{i2} - \varepsilon_{i1}. \quad (6)$$

La calidad de d_i como un estimado de t'_i depende del tamaño de la variación entre parcelas representado por ε_{ij} , como se discute a continuación. Trazando la función de distribución acumulada ordenando los valores de d_i en orden creciente y luego trazándolos de nuevo contra i/n muestra la proporción (eje vertical) de las n fincas que alcanzan un efecto de tratamiento de x o

menos (eje horizontal) (Figura 2a, Figura 2b). Se pueden dibujar los intervalos de confianza alrededor de esta curva utilizando el estadístico D de Kolmogorov-Smirnov, como se implementa en el paquete R 'SFSMISC' (Figura 2a, Figure 2b) (Maechler, 2014).

En lugar de estimar y dibujar la función de distribución acumulada, un enfoque alternativo es asumir un modelo específico (Figura 2c, Figura 2d). Si se asume que los valores d_i tienen una distribución Normal, entonces la curva es

$$p(x) = \Phi + \left(\frac{x-t}{s_t} \right), \quad (7)$$

donde Φ es la función de densidad acumulada para la distribución normal estándar, t la media de los valores d_i y s_t su desviación estándar. Se puede dibujar un intervalo de confianza alrededor de esta curva observando que si:

$$z = \frac{x-t}{s_t}. \quad (8)$$

Entonces

$$se(z) \approx \sqrt{\frac{1}{n} + \frac{(x-t)^2}{2s_t^2(n-1)}}, \quad (9)$$

siendo $\Phi(z \pm 2se(z))$ los límites de confianza aproximados del 95% (Figura 2c, Figura 2d).

Tanto la curva de distribución acumulada empírica como la basada en un modelo tienen ventajas y desventajas. La curva empírica tiene límites de confianza más altos, particularmente en los extremos; si se desea calcular los riesgos en el extremo superior o inferior de la escala, se necesita una gran n y muchas fincas. Por otra parte, los cálculos de la curva basada en el modelo y sus límites de confianza están condicionados a que el modelo sea apropiado, pues de otra manera estarían sesgados. El modelo de distribución Normal utilizado aquí se caracteriza por ser simétrico, mientras que la distribución de los efectos de tratamiento a lo largo de las fincas a menudo es sesgada, con unas cuantas fincas que muestran grandes efectos positivos. Si la distribución está sesgada entonces el modelo normal no será apropiado. Dibujar intervalos de confianza alrededor de las curvas de frecuencia enfatiza que se trata de estimados y que necesariamente implican cierto nivel de incertidumbre (Figura 2). También señala que con los tamaños de las muestras en estos experimentos (77 y 71 fincas para el maíz y el fríjol, respectivamente), se obtienen buenos estimados para el riesgo en la mitad del rango, pero el intervalo de confianza es amplio en la parte superior e inferior de la cola. Se necesita mucha observación para calcular los riesgos más extremos, con implicaciones para el diseño de tales experimentos, como se analiza más adelante. Los cálculos basados en el modelo ofrecen buena precisión en los extremos pero son sesgados, a menos que los supuestos del modelo sean realistas. Si el modelo utilizado no es ciertamente idóneo, entonces una opción sería encontrar un modelo alternativo que describa mejor la distribución de los datos.

Variación y riesgo

Las curvas de frecuencia acumulada para datos provenientes de los conjuntos de datos para el maíz y el frijol muestran que mientras los incrementos de la media son claros y positivos, hay gran variación entre las fincas en lo concerniente a rendimientos y al tamaño de los efectos de tratamiento. Presentar la información de esta manera pone de relieve los riesgos para los agricultores. El efecto de tratamiento en un experimento en una finca determinada es el mejor estimado que tenemos de lo que ocurriría con los rendimientos si el agricultor cambiara la práctica pasando del control al nuevo tratamiento. Para la base de datos del maíz, solo 6 de 77 (o el 8%) de las fincas tuvieron un incremento dentro del 10% del incremento de la media de 1.5 t ha^{-1} . En el 10% de las fincas se observó un aumento de más de 4 t ha^{-1} , mientras que en el 10% de las fincas se observó una disminución en los rendimientos con el nuevo tratamiento. Para la base de datos del frijol, las cifras fueron similares (6% de las fincas cayó dentro del 10% de la media, en el 14% se observó una disminución, mientras que en el 5% hubo un incremento de $>1 \text{ t ha}^{-1}$). La decisión de un agricultor de pasarse al nuevo tratamiento estaría basada en muchos criterios, pero un propósito prioritario para pruebas como esta es generar recomendaciones. Si el incremento de la media se reportara y usara como base para una recomendación, entonces la mayoría de los agricultores vería resultados diferentes. La variación en los efectos de tratamiento a lo largo de las fincas representa un riesgo o una oportunidad y no debería ser ignorada.

Relevancia de la variación dentro de la parcelas

Los diseños experimentales utilizados (y recomendados, como se verá a continuación) tienen una repetición por finca. Por lo tanto, en la ecuación (3) no se puede distinguir t'_{ij} de ε_{ij} . Nosotros interpretamos que d_i está representando el efecto de tratamiento tal como se experimenta en la finca i . No obstante, si ε_{ij} es grande d_i se debería principalmente a la variación dentro del mismo terreno. Los diseños experimentales tradicionales se replican dentro de un campo, de manera que se pueda estimar dentro de ese campo la varianza σ^2 de ε entre parcelas. Si un experimento con dos tratamientos debe poder generar un estimado de σ^2 , entonces se necesitan por lo menos 9 repeticiones, dando 8 grados de libertad para estimar la varianza σ^2 dentro del campo, el mínimo absoluto de una estimación útil (Mead, 1988). Esto no sería realista para los agricultores y de hecho no tiene sentido desde el punto de vista científico por dos razones: la primera es que los recursos que se gastan en repeticiones dentro de un campo generarían más información si se utilizaran para muestrear más de la variación entre campos, para así poder aprender más acerca de la variación en t_i y los factores que influyen en ella. Nosotros consideramos que 1 repetición en cada una de 70 fincas da mucha más información que 7 repeticiones en solo 10 fincas. La segunda razón es que muchas veces estos campos son pequeños, y las parcelas experimentales son relativamente grandes. Incluso si los campos son extensos, los investigadores y agricultores pueden tomar medidas para garantizar que las parcelas experimentales sean relativamente homogéneas antes de aplicar los tratamientos, ya que usualmente los agricultores entienden los patrones de variación dentro del campo. En ese caso, no sería frecuente una gran variación entre

parcelas del mismo tratamiento dentro de un campo. Podría haber excepciones, por ejemplo en la agricultura de tala y quema (Ortiz, 1995), o en el Sahel (Voortman y col., 2004), donde se puede observar una variabilidad substancial dentro de distancias cortas, por lo que se necesitaría usar parcelas relativamente grandes.

Si la variación en los efectos de tratamiento a lo largo de las fincas (t_i) es σ_t^2 entonces el mejor estimado de t_i es t'_{si} , el BLUP (iniciales en inglés para “mejor predicción lineal no sesgada”), Robinson (1991):

$$t'_{si} = \frac{\sigma_t^2}{\sigma_t^2 + \sigma^2} d_i. \quad (10)$$

Los estimados en la ecuación (10) son los estimados que nosotros usamos disminuidos por un factor que depende del tamaño de la variación dentro del campo. Por ejemplo, en nuestra prueba con el maíz el factor es 0.97, 0.89 y 0.78 para los coeficientes de variación (CV) dentro del campo de 10, 20 y 30% respectivamente. Por lo tanto, incluso con valores de CV más altos, la variación en el tratamiento es todavía significativa y nuestra interpretación es realista. Si la variación entre una parcela y otra dentro de un campo llegara a ser mucho mayor que eso, entonces el estimado d_i no da un buen estimado de t_i y el diseño no es apropiado para analizar la variación entre fincas en el efecto de tratamiento.

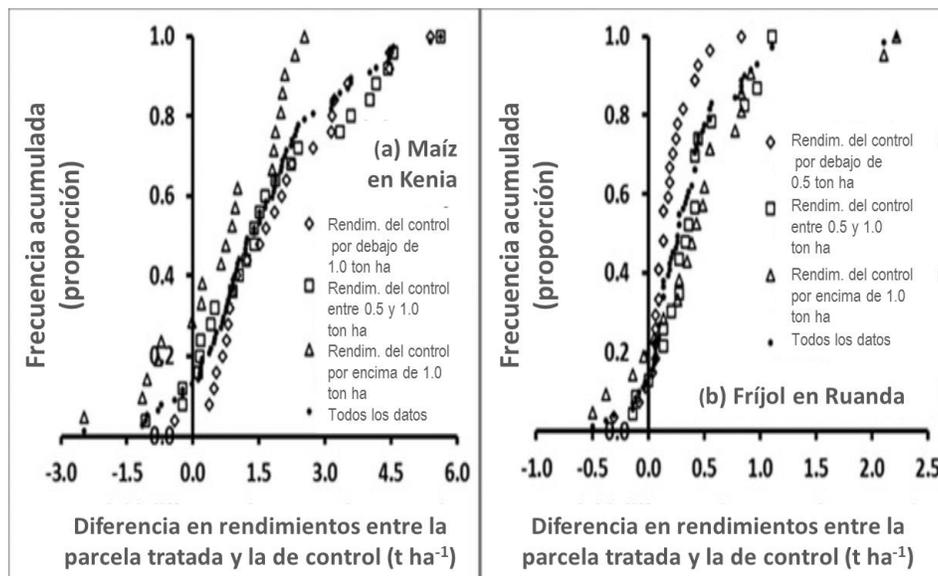


Figura 3. Curvas de frecuencia acumulada para todos los datos y para tres grupos de rendimientos de control para el conjunto de datos del maíz en Kenia (a) y el conjunto de datos del frijol en Ruanda (b).

Cómo explicar la variación

Con una variación tan fuerte entre finca y finca en el efecto de tratamiento, un objetivo del análisis de datos debería ser explicar la variación. Si hubiéramos medido las características relevantes de cada finca, entonces sería posible explorar aquellas que están relacionadas con t_i . Por lo tanto,

decidir qué características deberían medirse y escoger el diseño para que el enfoque sea lo más útil posible, debería ser una prioridad, como se discute más adelante. En los dos conjuntos de datos tenemos información limitada que se puede usar para explicar la variación. Una variable que es un indicador de las diferencias entre fincas es el rendimiento en las parcelas de control. Los datos se pueden dividir en grupos de rendimiento de control bajo, moderado y alto y la curva de riesgo estimada para cada grupo (Figura 3) (Biielders y Gérard, 2015; Ronner y col., 2016). Tanto el efecto medio o mediano del tratamiento como la varianza difieren entre estos grupos.

En lugar de una división en grupos arbitraria, la relación entre los desempeños de los dos tratamientos se puede examinar gráficamente, trazando y_{i2} contra y_{i1} . Esto es más fácil de interpretar cuando t_1 es un tratamiento de control o de línea base. Utilizando la ecuación (3):

$$y_{i2} = y_{i1} + t'_i + \varepsilon'_i , \quad (11)$$

donde $\varepsilon_i = \varepsilon_{i2} - \varepsilon_{i1}$ es un término de error. De ahí que, si no hubiera efecto de tratamiento, el gráfico sería una dispersión alrededor de la línea 1:1. Si hubiera un efecto de tratamiento t constante el gráfico mostraría una dispersión alrededor de la línea con la pendiente 1 y la intercepción t (Figura 4a). Una desviación común desde un efecto de tratamiento constante es que el tamaño del efecto de tratamiento dependa del rendimiento del tratamiento de control ya que este representa “la calidad de la parcela”. Por lo tanto, podríamos agrupar las unidades de observación por rendimiento del control, y encontrar el efecto medio para cada grupo o, menos arbitrariamente, poner una curva suave a través del trazo de la dispersión. En la Figura 4 se ilustran otras posibles formas de las curvas. La Figura 4d parece ser la más típica para los tratamientos de fertilidad del suelo, aunque la Figura 4b es lo que uno podría esperar que tuviera impacto sobre las parcelas que necesitan los mayores nutrientes. En la Figura 5 dibujamos una curva suave utilizando la regresión polinomial local tal como está implementada en R (R Equipo central, 2014). Las respuestas en nuestros datos (Figura 5) son más similares a la Figura 4d. Para el maíz, el efecto medio de tratamiento es pequeño y positivo en el extremo más bajo, más o menos constante hasta rendimientos de control de alrededor de 3 t ha⁻¹, y luego baja a cero. Para el fríjol, no hay efecto de tratamiento en las parcelas menos productivas. Incrementa gradualmente hasta que los rendimientos de control son alrededor de 2 t ha⁻¹ y luego baja a cero. En ambos casos se puede ver un efecto de tratamiento bajo o inexistente en las parcelas más pobres, como se observa comúnmente. Los suelos en este tipo de campos se conocen a menudo como “suelos no reactivos” (Vanlauwe y col., 2010). En los campos que muestran un mejor desempeño los rendimientos están alcanzando aquellos que se logran con la tecnología y no es de esperar que la adición de más nutrientes incremente el rendimiento.

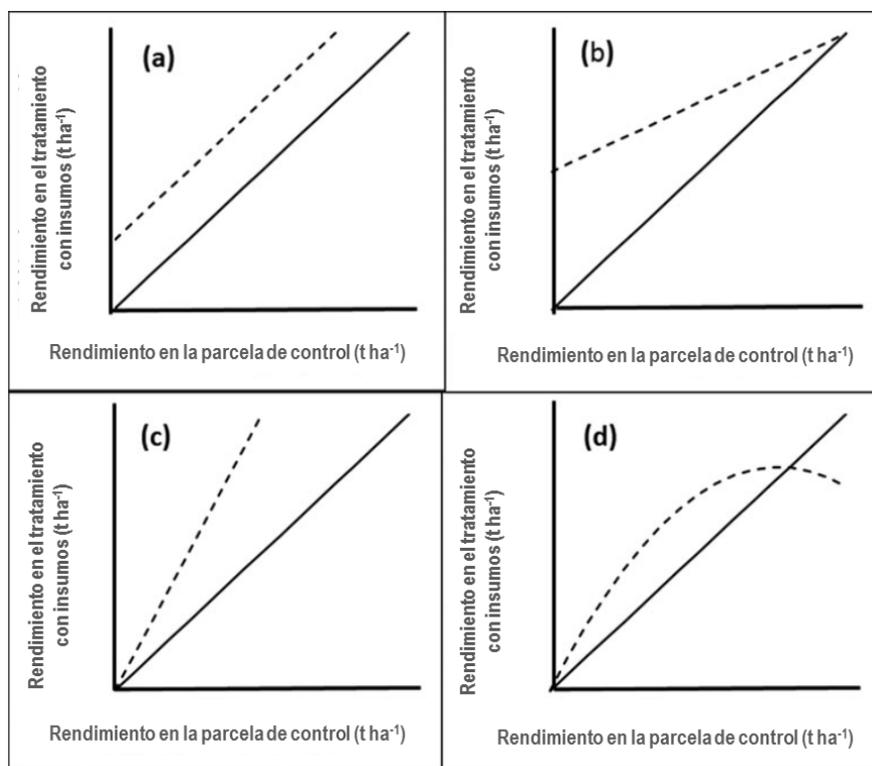


Figura 4. Respuestas posibles (líneas punteadas) cuando las unidades de observación en una parcela tratada (eje vertical) se trazan contra el control o línea base en la misma finca (eje horizontal). (a): efecto de tratamiento constante; (b): efecto de tratamiento positivo sobre las parcelas más pobres que disminuye con la calidad de la parcela; (c): efecto de tratamiento positivo que aumenta con la calidad de la parcela y (d): el efecto de tratamiento es cero en las parcelas más pobres, luego positivo, luego negativo en las mejores parcelas. La línea continua indica la línea 1:1.

El conjunto de datos de rendimiento del frijol de Ruanda tiene una fuente adicional de información que puede utilizarse para dar luces sobre la variación en el efecto de tratamiento a lo largo de las fincas, ya que el estudio se realizó en tres zonas diferentes. Tanto la media como la variación en el efecto de tratamiento y la probabilidad de observar un efecto negativo difieren entre las zonas (Tabla 2).

Tabla 2. Efectos de tratamiento (rendimiento con fosfato diamónico menos rendimiento de control) por distrito para el frijol en Ruanda.

Distrito	Efecto de tratamiento sobre el rendimiento ($t\ ha^{-1}$)		
	Promedio	Desviación estándar	Proporción de campos con respuesta negativa
Bugesera	97	270	0.32
Kibungo	266	259	0.00
Umutara	549	500	0.05
Total	358	457	0.14

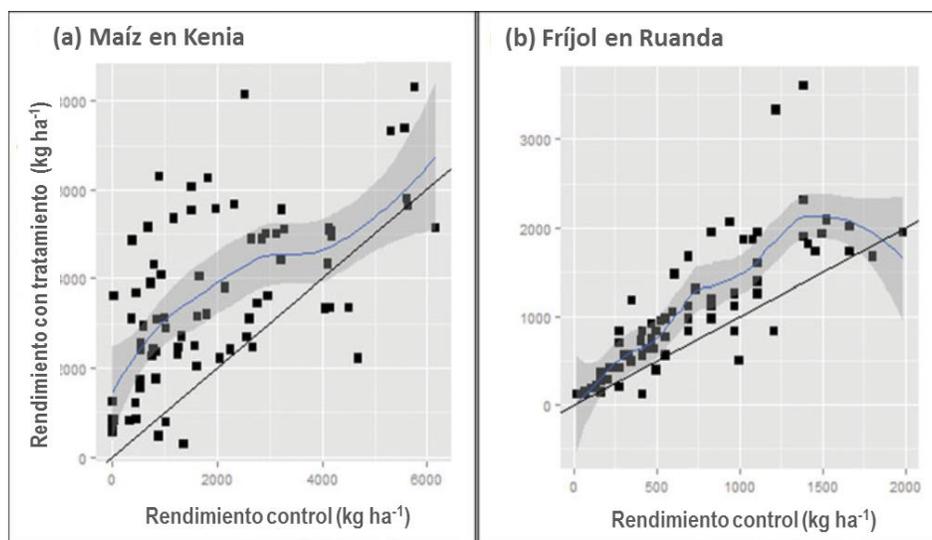


Figura 5. Gráficos de dispersión que representan el rendimiento en los tratamientos con aplicación de fertilizante frente a los rendimientos de control con adición de curvas de respuesta modeladas e intervalos de confianza para los conjuntos de datos del maíz de Kenia (a) y del fríjol de Ruanda.

DE MEDIR LA VARIACIÓN A GENERAR RECOMENDACIONES MATIZADAS

Cómo modificar las curvas de frecuencia acumulada

Las recomendaciones óptimas para los agricultores acerca de las opciones en el manejo de cultivos normalmente se basan en los resultados de pruebas analizadas y reportadas de la manera clásica, utilizando medias. Aunque se puede usar otra información, como por ejemplo evaluaciones económicas y de los agricultores, las evaluaciones cuantitativas básicas del valor de las nuevas prácticas se determinan usualmente con esa información. Pero con una variación tan grande entre fincas estas recomendaciones basadas en la media ofrecen una orientación equivocada. Pocos agricultores experimentan cambios cercanos a la media, unos beneficiándose mucho más y otros mucho menos. Un objetivo de la investigación debería ser cambiar la situación, para que las recomendaciones y los alcances pronosticados sean más confiables para más personas y/o tierras. Esto requiere explicar o comprender las fuentes de la variación que se muestran en la curva de riesgo, con los factores que contribuyen a una finca en particular situados hacia el extremo inferior o superior de la curva. Si un agricultor considerara la posibilidad de cambiar su manejo, pasando del que está representado por el control al que está representado por el tratamiento con una probabilidad conocida sobre dónde su efecto podría caer en la curva de frecuencia acumulada, entonces podría usar esa información por lo menos de dos maneras. La primera, si el factor está dentro del ámbito de control del agricultor (por ejemplo factores de manejo tales como el tiempo de siembra en relación con otros cultivos), él puede elegir utilizar niveles apropiados; y la segunda, si el factor está fuera del control del agricultor (por ejemplo el tipo de suelo) él puede optar por no hacer el cambio.

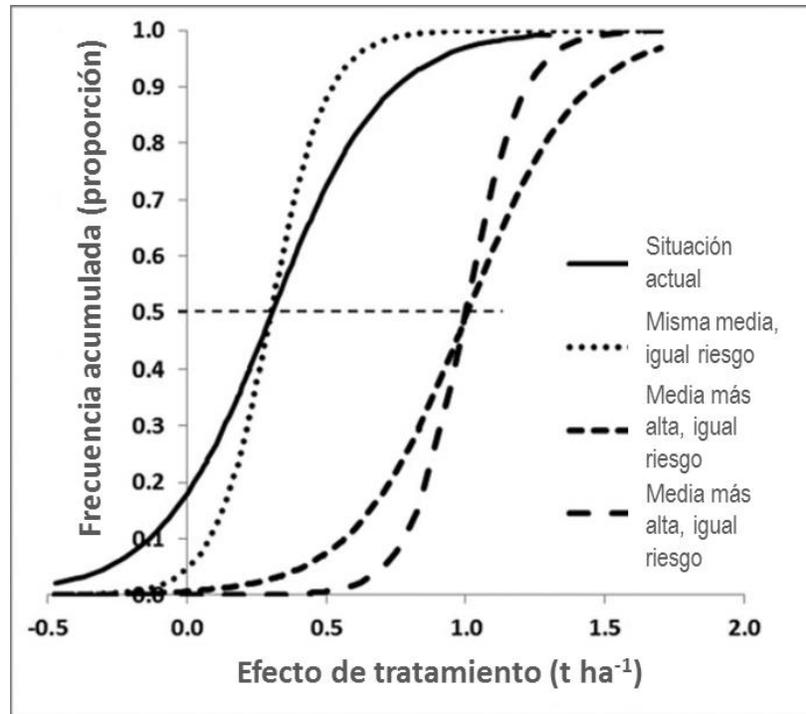


Figura 6. Las recomendaciones mejoradas podrían incluir mover la curva de frecuencia acumulada hacia la derecha (lo que resulta en efectos de tratamiento más altos para proporciones similares), crear una curva más vertical (lo que resulta en recomendaciones más predictivas), o una combinación de ambas. La línea punteada horizontal se intersecta con las curvas de frecuencia acumulada en valores x medios

Una comprensión refinada de cuáles son las causas para la variación observada resulta en curvas de frecuencia acumulada modificadas. El riesgo de un resultado pobre (por ejemplo un efecto negativo en la producción) se puede reducir desplazando la curva de frecuencia acumulada hacia la derecha, haciéndola más recta, o ambas (Figura 6). Desplazar la curva horizontalmente corresponde a cambiar la media, mientras que hacerla más recta es una disminución en la varianza. También sería posible cambiar la forma, por ejemplo reduciendo la probabilidad de una respuesta extremadamente baja, que corresponde a tirar de la cola inferior de la curva. En la práctica, la información refinada acerca de la práctica sustituirá la curva de riesgo única con varias curvas, cada una condicionada por factores ya sea ecológicos o de manejo. Los resultados de ambos estudios de caso proporcionan ejemplos (Figura 3). Ya que estas curvas generan diferente información y recomendaciones para campos con productividad actual baja, media y alta, una recomendación uniforme sería reemplazada por recomendaciones específicas según el tipo de campo.

Hacia un marco para aumentar la solidez de las recomendaciones

En esta sección y en la Figura 7 se presenta una estrategia para investigar el problema genérico descrito hasta ahora. El proceso es iterativo, con preguntas de investigación que van evolucionando y múltiples rondas de experimentación. Los investigadores plantean la hipótesis de que algunas nuevas opciones, por ejemplo las prácticas de manejo de la fertilidad del suelo, serán

de utilidad para los agricultores, y proceden a diseñar y llevar a cabo pruebas para investigar y demostrar su efecto. En la mayoría de los experimentos agronómicos las opciones consideradas definen los tratamientos del experimento. Como vemos a continuación, hay otros factores que también podrían determinar los tratamientos, así que acá usamos el término *opciones* para describir las prácticas nuevas o alternativas que se están estudiando. El punto de partida es definir el área y la población objetivos para los cuales se necesitan los resultados. Sabemos que las opciones que se investigan van a interactuar con el contexto biofísico, social y económico, a los cuales aquí nos referiremos colectivamente como el *contexto*. Es necesario definir los límites para esos contextos que son considerados por la investigación; de lo contrario, el problema del diseño es inmanejable ya que siempre hay otro contexto para investigar. El objetivo o foco se puede definir ya sea geográficamente (por ejemplo tierras entre 1.000 y 1.800 msnm en Kenia) o por referencia a una población (por ejemplo agricultores con menos de 2 ha de tierra que cultivan maíz como cultivo básico) dentro de un área definida geográficamente. Los objetivos de la prueba se fijan de la manera habitual, pero deberán incluir explorar, comprender o probar hipótesis acerca de la naturaleza de las interacciones entre las opciones y los contextos (interacciones de $O \times C$). Las hipótesis podrían referirse a respuestas biofísicas (tales como el rendimiento de los cultivos), manejo actual por parte de los agricultores, recursos de los que disponen los agricultores, y sus preferencias. Sea lo que sea que se mida, los problemas de diseño experimental son básicamente los mismos. Si no se planteó ninguna hipótesis sobre las interacciones de $O \times C$, entonces la prueba podría continuar con un diseño simple en el que cada agricultor tenga una repetición de cada tratamiento. La falta de interacciones hipotéticas (o de las que se tenga conocimiento previo) significa que no hace ninguna diferencia cuáles fincas o agricultores se incluyan. Pueden ser una muestra aleatoria o, más práctico y más común, voluntarios de grupos de agricultores con quienes trabajan las organizaciones que hacen labores de extensión.

Si se espera que las interacciones de $O \times C$ sean importantes, las posibilidades de diseño experimental dependen de la naturaleza de los factores que se supone que interactúan con esas opciones. Estos son de tres tipos principales, con diferentes posibilidades de diseño (Figura 7). En primer lugar, los factores que se pueden mapear son los definidos por una posición geográfica, de modo que es posible señalar una ubicación en un mapa y predecir con alta precisión cuál es el valor del factor en esa ubicación. Un ejemplo es la zona agroecológica. Estos factores se utilizan para definir los lugares para la prueba. Un factor como el estado de P del suelo puede en principio ser mapeable, y se están produciendo mapas de alta resolución de las propiedades del suelo (www.africasoils.net). Sin embargo, es poco probable que estos mapas sean lo suficientemente precisos como para ayudar a escoger lugares a escala fina (van Apeldoorn y col., 2014). En segundo lugar, los factores predecibles pero no mapeables son aquellos que son predecibles en el sentido de que no cambian (al menos durante la escala de tiempo de un experimento típico) pero no pueden predecirse a partir de un mapa. El diseño experimental utiliza la estratificación para hacer frente a estos factores. Los factores predecibles pueden variar entre diferentes fincas y agricultores, como es el caso del género y el nivel de recursos de los agricultores. El diseño podría entonces incluir deliberadamente a agricultores de recursos bajos, medios y altos, o a cada género

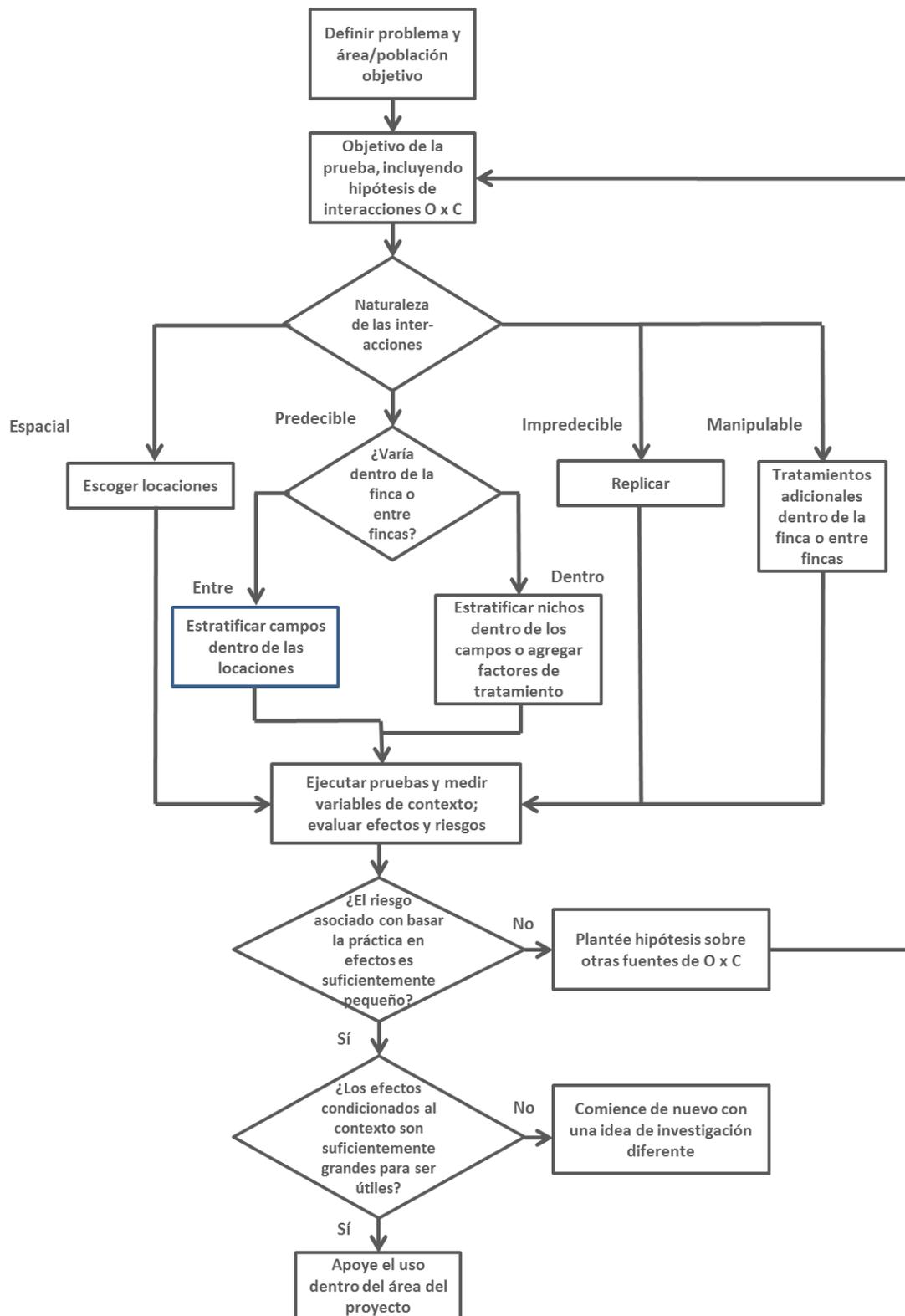


Figura 7. Modelo propuesto para integrar la heterogeneidad dentro de las iniciativas que buscan implementar para cada sitio en particular recomendaciones de manejo de fertilidad del suelo para los pequeños productores.

'O x C' significa interacciones de Opciones x Contexto

Pero también podría incluir factores que varían al interior de una finca, por ejemplo, campos más cerca o más lejos de la vivienda, o los campos clasificados por los agricultores como pobres o con suelo más fértil. El diseño de la prueba podría entonces incluir comparaciones de estratos dentro de las fincas, las cuales por lo general serán más precisas que entre las diferentes fincas. También podría ser más fácil de organizar, ya que los investigadores no tendrían que negociar con tantos agricultores. En tercer lugar, los factores impredecibles son aquellos para los cuales no se conoce su nivel hasta después de realizado el experimento. Algunos ejemplos son el clima, los brotes de plagas o la calidad de manejo de la parcela. La única opción de diseño aquí es replicar en espacio y tiempo, para que haya suficientes fincas para tomar una muestra de la variación.

También existen factores de contexto que hipotéticamente influyen en los resultados, que podrían convertirse en factores de tratamiento ya que pueden manipularse y asignarse aleatoriamente a las parcelas (Figura 7). Un ejemplo es la fecha de siembra. Si algunos agricultores siembran temprano y algunos tarde en una temporada de crecimiento con lluvias normales, entonces esto podría ser incluido como un factor "predecible", con algunos agricultores que acuerdan plantar temprano y otros tarde. Pero sería más eficiente incluirlo como un tratamiento aleatorio, al menos en algunas de las fincas. Si un factor como la fecha de siembra no se considera durante el diseño y se deja a elección de los agricultores, será un factor "impredecible". Si el factor sí interactúa con los efectos de tratamiento (es decir, la diferencia en el desempeño de los dos tratamientos depende de la fecha de siembra), entonces la variabilidad en las fechas de siembra contribuirá a la incertidumbre de los resultados. Si se prevé que la influencia del factor es importante, entonces se puede cambiar a otra categoría. Se obtendrá la mejor precisión cuando se incluya el factor como factor de tratamiento, de modo que las parcelas se siembren temprano o tarde deliberadamente (con las fechas de siembra definidas a conveniencia). Un factor como este podría incluirse en el conjunto de tratamientos dentro de una finca o campo (con los beneficios de precisión y un conjunto de resultados más significativo en cada finca) o a lo largo de diferentes fincas (con los beneficios de mantener las pruebas más simples y pequeñas para un agricultor individual). Tenga en cuenta que no se tiene que tomar la misma decisión para todas las fincas; si algunos agricultores estuvieran interesados en un conjunto de tratamientos más grande, entonces podrían incluir fechas de siembra diferentes, mientras que otros usan una sola fecha de siembra.

El número total de fincas necesarias para la prueba dependerá del número de categorías en la lista de factores de contexto que se van a investigar. Una regla general podría ser incluir al menos 20 fincas de cada categoría, pues este tamaño de muestra dará buena información sobre el resto de la variabilidad entre fincas dentro de cada categoría de contexto. El principal punto en que la ejecución de este tipo de pruebas difiere de la norma es en que los factores de contexto que hipotéticamente interactúan en la determinación de los alcances deben ser medidos y registrados (Figura 7). Estos incluyen todos los factores planteados como hipótesis en la etapa de diseño y desarrollados en el diseño de los tres tipos anteriores. Existe la tentación de medir más variables que "podrían ser interesantes", pero este es un enfoque defectuoso. Si usted tiene una fuerte intuición de que hay otras variables de contexto que podrían ser importantes en las interacciones, entonces inclúyalas en el proceso de diseño que se muestra arriba, pues de lo contrario usted

podría no terminar con un diseño eficiente para comprender su efecto. Si realmente no hay fundamento para que estas variables interactúen con las opciones, entonces medirlas es desperdiciar esfuerzos. Puede haber otros factores que aparecen durante la temporada. Si la inspección de las parcelas en estudio durante la temporada sugiere que hay una variación en los efectos de tratamiento entre fincas que se esperaría fueran similares, entonces los investigadores deben (a) formular hipótesis sobre las razones de esto y (b) agregar variables a las mediciones planificadas, para que las hipótesis sean examinadas en el momento del análisis. Se pueden utilizar los principios de diseño de tratamientos para mejorar la eficacia de los diseños. Algunos ejemplos son el uso de combinaciones factoriales de factores de contexto y elección de fracciones adecuadas de ellos, así como directrices para elegir niveles experimentales de factores continuos. Inevitablemente habrá complejidades y cuestiones de diseño que son difíciles de resolver: por ejemplo, cómo tratar con factores de contexto que están altamente correlacionados, tales como diferentes propiedades del suelo. Pero conocer los principios de diseño y las formas de abordar los factores de contexto de los diferentes tipos descritos aquí puede conducir a diseños experimentales que son mucho más eficientes y efectivos que la selección arbitraria o aleatoria de las fincas y parcelas a incluir.

El análisis de los resultados debe incluir la estimación del tamaño de los efectos de tratamiento para cada combinación de factores de contexto sobre los que se han formulado hipótesis y cuya importancia se ha confirmado. Debe incluirse también la variación no explicada entre todos los campos o fincas (Figura 7). Con cualquiera de los dos enfoques, el riesgo o incertidumbre es la variación residual o no explicada entre campos o fincas en el efecto de tratamiento. Hay dos alcances para el cálculo de este riesgo, expresados como respuestas a la pregunta: "¿Es el riesgo o la variación no explicada entre campo y campo o finca y finca en el efecto de tratamiento suficientemente pequeño?" Si la respuesta es "sí", entonces conocer el contexto da una buena predicción del efecto de tratamiento. En este caso, la interpretación se puede hacer según las líneas habituales. Eso significa que nosotros decidimos si esos efectos de tratamiento, que están condicionados por el contexto, son lo suficientemente grandes como para ser "útiles", es decir, para significar una diferencia para los agricultores, ser económicamente viables, imponerse sobre otras limitaciones, etc., y luego podemos ponerlos a disposición para su difusión. Si la respuesta es "no", quiere decir que la variación entre campos o fincas en el efecto de tratamiento es todavía grande, de modo que la información sobre el efecto de cualquier tratamiento en un campo o finca individual es demasiado incierta para servir de base a recomendaciones y decisiones. En este caso, es necesario plantear la hipótesis de otros factores responsables de la variación e iniciar una nueva ronda de experimentación.

Las ideas acerca de cuáles podrían ser los factores pueden provenir del conocimiento científico y las observaciones de los investigadores durante la temporada, o del conocimiento de los agricultores. Por ejemplo, un proyecto de manejo de la fertilidad del suelo en Malawi encontró una gran variación entre una finca y otra en los efectos que tuvo el cultivo intercalado de leguminosas sobre los rendimientos de maíz. Los científicos tenían la hipótesis de que los responsables eran los factores del suelo, particularmente COS y P. El proyecto también utilizó un

método de *análisis participativo de varianza* (PANOVA) y los agricultores sugirieron que la principal fuente de variación fue la calidad del manejo de la parcela (preparación de la tierra, tiempo de siembra, tiempo de deshierbe y minuciosidad). Estos factores se incluyeron en la siguiente temporada: las variables del suelo como "predecibles" y la calidad del manejo de la parcela como un factor impredecible que se midió con un cuestionario estándar simple (WMhango, comunicación personal). En nuestro ejemplo con el fríjol, los agricultores, técnicos o investigadores podrían tener ideas sobre las razones detrás de los dos valores atípicos en la [Figura 5b](#) los cuales tienen respuestas inusualmente buenas frente al tratamiento.

La exploración de datos a menudo revela efectos de localización, como en nuestro conjunto de datos del frijol ([Tabla 2](#)). Entender la fuente de los efectos de localización debería ser otra ruta para generar hipótesis de $O \times C$. En nuestro caso, la respuesta media en Bugesera fue débil, pero la respuesta en algunas fincas en esa ubicación fue buena. Entonces, se requiere información sobre los factores que conducen a una respuesta débil y a una respuesta buena en Bugesera, así como las explicaciones para las diferencias entre unos lugares y otros. La conclusión de que "en el lugar x el efecto es diferente al del lugar y " no es satisfactoria, ya que no da ninguna pista sobre lo que podría suceder en otros lugares. Las hipótesis de los mecanismos detrás de los efectos de lugar tienen que ser desarrolladas y probadas con diseños apropiados. Si los factores hipotéticos sólo varían entre "lugares", por ejemplo la duración del período de crecimiento, entonces será necesario incluir en el diseño lugares adicionales.

En el manejo del suelo es común depender del camino o los patrones trazados: las respuestas en una temporada dependen del manejo y de las respuestas a ese manejo en temporadas anteriores. Si se considera que el manejo previo es una fuente importante de variación, entonces es necesario formular y probar hipótesis. Estas pueden ser en términos de prácticas de manejo en el pasado (por ejemplo, distinguir si el último sistema de cultivo fue el mismo o diferente de esta temporada, o el tiempo transcurrido desde que la finca fue convertida de bosque), condiciones mensurables del suelo (por ejemplo, el contenido de carbono orgánico en el suelo), los niveles históricos de producción, o tal vez la clasificación de los suelos por parte de los agricultores. Los principios de diseño de la prueba para probar tales hipótesis son los mismos en cada caso.

Una forma de analizar esa variación tan alta en las respuestas a los tratamientos experimentales con el fin de obtener ideas que lleven a hipótesis es usar líneas límite ajustadas a los valores máximos de respuesta a través del rango de rendimientos para un rango de factores determinantes del rendimiento (por ejemplo Fermont y *col.*, 2009; van Asten y *col.*, 2003). La línea de límite representa la mayor respuesta esperada para un valor dado de un factor determinante del rendimiento, por ejemplo la disponibilidad de cierto nutriente, y se asume que los puntos que caen por debajo de la línea de límite están limitados por uno de los otros factores medidos, por ejemplo, la disponibilidad de otro nutriente o la incidencia de una plaga o enfermedad importante. Al comparar las líneas de límites trazadas cuando los rendimientos se representan frente a una serie de factores, se puede estimar la contribución relativa de estos factores al intervalo de rendimiento. Los métodos apropiados de análisis de datos dependen de los detalles del diseño, las

variables de contexto medidas y las hipótesis de interacciones. En el caso más simple, con dos tratamientos y todas las variables de contexto medidas a nivel de finca, el análisis es más fácil si se basa en diferencias (Ecuación 4). El objetivo de los métodos estadísticos es relacionar las diferencias entre d_i y las variables de contexto medidas. Los modelos de regresión, lineales o no lineales según corresponda, son la herramienta principal. Si se dispone de grandes conjuntos de datos (número de fincas), se pueden utilizar métodos de descubrimiento de patrones tales como bosques aleatorios (Breiman, 2001). Comparados con los métodos de regresión más tradicionales, estos tienen la ventaja de que identifican y describen interacciones complejas, mientras que la desventaja es que no están basadas en la comprensión o las hipótesis del proceso y que requieren muestras grandes.

Consideraciones prácticas

El gran número de sitios de prueba o parcelas agrícolas que se necesitan para llevar a cabo los experimentos descritos anteriormente significa que los procedimientos experimentales deben ser simples y con una mayor participación de los agricultores en algunos puntos que constituyen un cuello de botella en los experimentos, tales como el manejo de la prueba y la recolección de datos. Con relación al manejo de la prueba, considerando que los componentes específicos de manejo pueden requerir un nivel mínimo de estandarización, alineado con la estructura de tratamiento de las pruebas, es posible que sea necesario contar con estructuras de capacitación y recolección de datos que operen entre el investigador y la familia campesina participante. Es frecuente que estas estructuras sean facilitadas a través de organizaciones no gubernamentales (ONG) que trabajan con grupos de agricultores, cooperativas, clubes de ahorro u otras organizaciones sociales que son autoelegidas o que se forman alrededor de un objetivo en particular. El hecho de que las ONG faciliten la captura de datos científicos con la calidad requerida, a menudo plantea un obstáculo práctico importante en la implementación de la estrategia de la Figura 7. Dicho esto, estas pruebas multilocacionales con un gran número de repeticiones se han implementado antes (Giller y col., 2013; Ronner y col., 2016) y la Figura 7 busca incrementar la eficiencia de tales pruebas en el futuro.

Si bien es posible que tengamos las herramientas analíticas, garantizar la recolección de datos de calidad es uno de los grandes problemas de las pruebas multilocacionales que son implementadas por los aliados del desarrollo. No deben subestimarse los desafíos logísticos que supone el tener un gran número de sitios de prueba diseminados a través de amplias áreas geográficas, la cantidad de operadores involucrados y las dificultades de estandarizar medidas, registrar e ingresar. Es preciso optar por una elección estratégica para medir un número limitado de variables basadas en preguntas e hipótesis claras, en lugar de recopilar datos por los datos en sí.

El gran número de sitios de prueba también favorece un modelo construido alrededor de organizaciones sociales existentes que a menudo se forman en torno a objetivos específicos que excluyen la representación de la comunidad agrícola en general. Por ejemplo, en el estudio de Franke y col. (este volumen), un grupo de agricultores recibió apoyo de una ONG dedicada

específicamente a las mujeres agricultoras desfavorecidas. Esto resultó en que un conjunto de campos con suelos muy poco fértiles era representativo de ese grupo de agricultores, pero probablemente no era representativo de la población más amplia de las fincas en esa aldea como un todo. Dado que las opciones específicas, así como el contexto, probablemente interactúan con las dotaciones de recursos y los objetivos de producción de los agricultores (Tittonell y col., 2010) y con las actitudes hacia la agricultura (www.tnsglobal.com), considerar el tipo de familia campesina con la cual comprometerse es un componente crítico al definir la población objetivo (ver arriba). Las limitaciones de las estructuras sociales existentes en términos de inclusión social pueden por lo tanto interponer una limitante práctica para la implementación de lo que sería una estrategia teóricamente sólida. Nelson y col. (este volumen) proponen un enfoque de "red de agricultores investigadores" para hacer viable la experimentación a gran escala.

CONCLUSIONES

La dependencia de los investigadores en una gama limitada de diseños experimentales formales se ha convertido en una camisa de fuerza cuando se trata de entender las respuestas en los campos de los agricultores. A pesar de nuestra mejor comprensión de la complejidad de los sistemas agrícolas de pequeños productores y del hecho de que el *Libro Verde* (Patel y col., 2004) haya existido desde hace una década, la próxima generación de científicos agrícolas sigue siendo entrenada en un modo clásico para tratar de reducir la variabilidad y centrarse en factores individuales. En este trabajo, exploramos una serie de enfoques para permitir que la investigación agronómica pueda explorar la gama completa de factores que determinan la productividad de los cultivos en el mundo real.

Hoy en día es plenamente reconocido que los agricultores deben estar integrados en el desarrollo y validación de opciones agronómicas mejoradas. Es reconocido también que los entornos agrícolas de pequeños productores son notoriamente heterogéneos, como lo demuestra la gran variación en las condiciones de fertilidad del suelo, y que las fincas y los agricultores tienen todas diferentes características, entonces no pueden ser considerados utilitariamente como "réplicas" de nada. Las curvas de frecuencia acumulada basadas en pruebas llevadas a cabo en un número relativamente grande de fincas permiten interpretar el riesgo asociado con una opción agronómica alternativa específica. Los conjuntos de datos de estudios de casos mostraron que sólo una pequeña proporción de campos generó resultados que están dentro del 10% de los efectos medios. Se propone una estrategia basada en las interacciones anticipadas entre las opciones de agronomía alternativas y el contexto de la agricultura, con interacciones espaciales "mapeables", predecibles "no mapeables" e impredecibles que conducen (multitemporada) a un proceso R4D basado en estratificación, estructuras de tratamiento modificadas y observaciones prioritarias, que finalmente resultan en recomendaciones con efectos de tratamiento y riesgos asociados aceptables. Las consideraciones prácticas deberán guiar la implementación de este marco teórico.

Si bien ofrecemos sugerencias sobre cómo avanzar en la comprensión de la complejidad de las respuestas de los cultivos a las intervenciones en fincas de pequeños productores, no pretendemos haber llegado a una comprensión perfecta de cómo abordar todos estos problemas. Esto deja un nutrido conjunto de nuevas preguntas para guiar la investigación en fincas en el futuro. Con el paso hacia la investigación agronómica en la finca, el asunto de variación y variabilidad en las respuestas ocupa un lugar central. En lugar de evitar la diversidad de respuestas a los tratamientos deberíamos acogerla, para así lograr un entendimiento más completo del contexto actual y su dependencia del manejo pasado.

Reconocimientos. Agradecemos a la Fundación Rockefeller y más específicamente al Dr. John Lynam, quien fuera uno de los primeros en invertir en una iniciativa orientada a descifrar la heterogeneidad en los sistemas de producción de pequeños agricultores. Expresamos también nuestro reconocimiento a la Dirección General de Desarrollo de Bélgica y a la Fundación Bill & Melinda Gates por su apoyo a nuestro trabajo sobre cómo adaptar la gestión de la fertilidad del suelo a las condiciones de los pequeños productores.

REFERENCIAS

- Anderson, G. D. (1974). Bean responses to fertilizers on Mt. Kilimanjaro in relation to soil and climatic conditions. *East African Agricultural and Forestry Journal* 39(3):272–288.
- Bielders, C. L. and Gérard, B. (2015). Millet response to microdose fertilization in south-western Niger: Effect of antecedent fertility management and environmental factors. *Field Crops Research* 171:165–175.
- Birner, R., Davis, K., Pender, J., Nkonya, E., Anandajayasekeram, P., Ekboir, J., Mbabu, A., Spielman, D. J., Horna, D., Benin, S. and Cohen, M. (2009). From best practice to best fit: A framework for designing and analyzing pluralistic agricultural advisory services worldwide. *The Journal of Agricultural Education and Extension* 15:341–355.
- Breiman, L. (2001). Random forests. *Machine Learning* 45:5–32.
- Buerkert, A., Bationo, A. and Piepho, H. (2001). Efficient phosphorus application strategies for increased crop production in sub-Saharan West Africa. *Field Crops Research* 72:1–15.
- Carter, J. (1995). Alley farming: have resource-poor farmers benefited? *Natural Resources Perspectives* 3:1–8.
- Cochran, W. G. and Cox, G. M. (1957). *Experimental Designs*. London: Wiley.
- Collinson, M. (2000). *A History of Farming Systems Research*. p. 448. Wallingford: CABI.
- Djurfeldt, G., Holmén, H., Jirström, M. and Larsson, R. (2005). *The African Food Crisis. Lessons from the Asian Green Revolution*. Wallingford: CABI Publishing.
- Fermont, A. M., van Asten, P. J. A., Tittonell, P., van Wijk, M. T. and Giller, K. E. (2009). Closing the cassava yield gap: An analysis from smallholder farms in East Africa. *Field Crops Research* 112:24–36.
- Fertilizer Use Recommendation Project (FURP) (1994). *Fertilizer Use Recommendations, Volumes 1–24*. Nairobi: Kenya Agricultural Research Institute.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.

- Giller, K. E., Franke, A. C., Abaidoo, R., Baijukya, R., Bala, A., Boahen, S., Dashiell, K., Kantengwa, S., Sanginga, J.M., Sanginga, N., Simmons, A., Turner, A., de Wolf, J., Woomeer, P. and Vanlauwe, B. (2013). N2Africa: Putting nitrogen fixation to work for smallholder farmers in Africa. In *Agro-ecological Intensification of Agricultural Systems in the African Highlands*, 156–174 (Eds B. Vanlauwe, P. Van Asten and G Blomme). London: Earthscan.
- Giller, K. E., Tittonell, P., Rufino, M. C., van Wijk, M. T., Zingore, S., Mapfumo, P., Adjei-Nsiah, S., Herrero, M., Chikowo, R., Corbeels, M., Rowe, E. C., Baijukya, F., Mwijage, A., Smith, J., Yeboah, E., van der Burg, W. J., Sanogo, O. M., Misiko, M., de Ridder, N., Karanja, S., Kaizzi, C., K'ungu, J., Mwale, M., Nwaga, D., Pacini, C. and Vanlauwe, B. (2011). Communicating complexity: Integrated assessment of trade-offs concerning soil fertility management within African farming systems to support innovation and development. *Agricultural Systems* 104:191–203.
- Le Mare, P. H. (1968). Experiments on the effects of phosphate applied to a Buganda soil: III. A chemical study of the soil phosphate, the fate of fertilizer phosphate and the relationship with iron and aluminium. *The Journal of Agricultural Science* 70:281–285.
- Maat, H. and Glover, D. (2012). Chapter 8. Alternative configurations of agronomic experimentation. *Contested Agronomy: Agricultural Research in a Changing World*, 131–145 (Eds J. Sumberg and J. Thompson). London: Earthscan.
- Maechler, M. (2014). sfsmisc: Utilities from seminar fuer Statistik ETH Zurich. *R Package Version 1.0–26*. <http://CRAN.R-project.org/package=sfsmisc>.
- Mead, R. (1988). *The Design of Experiments: Statistical Principles for Practical Applications*. New York: Cambridge University Press.
- Nelson, R. and Coe, R. (2014). Transforming research and development practice to support agroecological intensification of smallholder farming. *Journal of International Affairs* 67(2):107–127.
- Ojiem, J. O., de Ridder, N., Vanlauwe, B. and Giller, K. E. (2006). Socio-ecological niche: A conceptual framework for integration of legumes in smallholder farming systems. *International Journal of Agricultural Sustainability* 4:79–93.
- Ortiz, R. (1995). Plot Techniques for assessment of bunch weight in banana trials under two systems of crop management. *Agronomy Journal* 87:63–69.
- Patel, B. K., Muir-Leresche, K., Coe, R. and Hainsworth, S. D. (2004). *The Green Book: A Guide to Effective Graduate Research in African Agriculture, Environment and Rural Development*, 1–250. Kampala: The African Crop Science Society.
- Poulton, P. R. (2006). *Rothamsted - The Classical Experiments*. Harpenden, UK: Lawes Agricultural Trust. R Core Team (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL: <http://www.R-project.org/>.
- Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science* 6:15–32.
- Ronner, E., Franke, A. C., Vanlauwe, B., Dianda, M., Edeh, E., Ukem, B., Bala, A., van Heerwaarden, J. and Giller, K. E. (2016). Understanding variability in soybean yield and response to P-fertilizer and rhizobium inoculants on farmers' fields in northern Nigeria. *Field Crops Research*, 186:133–145.
- Scaife, M. A. (1968). Maize fertilizer experiments in Western Tanzania. *The Journal of Agricultural Science*, 70:209–222.
- Sperling, L. and Ashby, J. A. (2000). Moving participatory breeding forward: the next steps. In *A History of Farming Systems Research*, 354–363 (Ed M. Collinson). Wallingford: CAB International.
- Sumberg, J. (2005). Constraints to the adoption of agricultural innovations - Is it time for a re-think? *Outlook on Agriculture* 34(1):7–10.
- Tittonell, P., Muriuki, A., Shepherd, K. D., Mugendi, D., Kaizzi, K. C., Okeyo, J., Verchot, L., Coe, R. and Vanlauwe, B. (2010). The diversity of rural livelihoods and their influence on soil fertility in agricultural systems of East Africa - A typology of smallholder farms. *Agricultural Systems* 103:83–97.

Tittonell, P., Vanlauwe, B., Leffelaar, P. A., Shepherd, K. D. and Giller, K. E. (2005) Exploring diversity in soil fertility management of smallholder farms in western Kenya. II. Within farm variability in resource allocation, nutrient flows and soil fertility status. *Agriculture, Ecosystems and Environment* 110:166–184.

Tittonell, P., Vanlauwe, B., Misiko, M. and Giller, K. E. (2011). Targeting resources within diverse, heterogeneous and dynamic farming systems: Towards a 'Uniquely African green revolution'. In *Innovations as Key to the Green Revolution in Africa: Exploring the Scientific Facts*, 747–758 (Eds A. Bationo, B. Waswa, J. M. Okeyo, F. Maina and J. Kihara). Dordrecht: Springer.

van Apeldoorn, D. F., Kempen, B., Bartholomeus, H. M., Rusinamhodzi, L., Zingore, S., Sonneveld, M. P. W., Kok, K. and Giller, K. E. (2014). Analysing soil organic C gradients in a smallholder farming village of East Zimbabwe. *Geoderma Regional* 2–3:32–40.

Van Asten, P., Wopereis, M., Haefele, S., Isselmou, M. O. and Kropff, M. (2003). Explaining yield gaps on farmer-identified degraded and non-degraded soils in a Sahelian irrigated rice scheme. *NJAS-Wageningen Journal of Life Sciences* 50(3-4):277–296.

Vanlauwe, B., Bationo, A., Chianu, J., Giller, K. E., Merckx, R., Mokwunye, U., Ohiokpehai, O., Pypers, P., Tabo, R., Shepherd, K., Smaling, R., Woomer, P. L. and Sanginga, N. (2010). Integrated soil fertility management: Operational definition and consequences for implementation and dissemination. *Outlook on Agriculture* 39:17–24.

Vanlauwe, B., Descheemaeker, K., Giller, K. E., Huisin, J., Merckx, R., Nziguheba, G., Wendt, J. and Zingore, S. (2015). Integrated soil fertility management in sub-Saharan Africa: Unravelling local adaptation. *SOIL* 1(1):491–508.

Vanlauwe, B., Tittonell, P. and Mukalama, J. (2006). Within-farm soil fertility gradients affect response of maize to fertilizer application in western Kenya. *Nutrient Cycling in Agroecosystems* 76:171–182.

Voortman, R. L., Brouwer, J. and Albersen, P. J. (2004). Characterization of spatial soil variability and its effect on Millet yield on Sudano-Sahelian coversands in SW Niger. *Geoderma* 121:65–82.

Waddington, S. R., Murwira, H. K., Kumwenda, J. D. T., Hikwa, D. and Tagwira, F. (Eds) (1998). *Soil Fertility Research for Maize-based Farming Systems in Malawi and Zimbabwe*. Harare, Zimbabwe: SoilFertNet/CIMMYT-Zimbabwe.

Zingore, S., Murwira, H. K., Delve, R. J. and Giller, K. E. (2007). Soil type, historical management and current resource allocation: Three dimensions regulating variability of maize yields and nutrient use efficiencies on African smallholder farms. *Field Crops Research* 101:296–305.