

Nuevo algoritmo para generar patrones espaciales de muestreo de suelos en experimentos agrícolas

New algorithm to compute soil spatial sampling patterns in agricultural experiments

 Edwin Erazo-Mesa^{1*}  Andrés Echeverri-Sánchez^{1*}  Norberto Urrutia Cobo¹

¹Universidad del Valle, Cali, Colombia

*Autor de correspondencia: Universidad del Valle, Ciudad Universitaria Meléndez Calle 13 # 100-00, Edificio E39, Oficina 2018A, Cali, Colombia.
osvaldo.erazo@correounivalle.edu.co

Recibido: 19 de octubre de 2020
Aprobado: 21 de julio de 2021
Publicado: 06 de diciembre de 2021

Editor temático: Judith del Carmen Martín Atencia, (Corporación Colombiana de Investigación Agropecuaria [AGROSAVIA])

Para citar este artículo: Erazo-Mesa, O. E., Echeverri-Sánchez, A.F., & Urrutia Cobo, N. (2022). Nuevo algoritmo para generar patrones espaciales de muestreo de suelos en experimentos agrícolas. *Ciencia y Tecnología Agropecuaria*, 23(1), e2306. https://doi.org/10.21930/rcta.vol23_num1_art:2306

Resumen: La variabilidad espacial del suelo es un factor importante para entender los cambios de las variables de respuesta en experimentos agrícolas. El muestreo de suelos se ejecuta con base en un patrón espacial, el cual puede ser aleatorio o sistemático. El objetivo de este trabajo fue validar un nuevo algoritmo para generar patrones espaciales de muestreo de suelos en este contexto. Para esto se diseñaron tres funciones en el software R, las cuales fueron comparadas con cinco aplicaciones (tres programas y dos librerías de R). La validación se realizó replicando tres patrones espaciales de suelos en experimentos agrícolas reportados en investigaciones anteriores, además de comparar la localización manual de puntos de muestreo en un experimento de cosecha de caña de azúcar con la localización generada por el algoritmo. Los resultados indican que el algoritmo tiene la capacidad exclusiva de realizar muestreos sistemáticos por unidad de área y centrarlos en el polígono correspondiente. El resto de las características, tales como el cálculo de los demás patrones y la generación de puntos sobre líneas, es posible encontrarlas en las otras aplicaciones. Con respecto a la validación en campo, la distancia promedio entre puntos generados con el algoritmo y los ubicados manualmente en campo es 2,58 m. La distancia promedio entre los puntos ubicados manualmente en campo y la línea de surco más cercana es 0,46 m. En conclusión, el algoritmo permite ubicar puntos de muestreo en sitios específicos del campo, como lo son las partes altas del surco o el entresurco.

Palabras clave: agricultura de precisión, cinemática en tiempo real, muestreo de suelo, software R, surco, variabilidad espacial

Abstract: Soil spatial variability is an essential factor in understanding the change of the dependent variables in agricultural experiments. Soil sampling is carried out based on a spatial pattern, which can be random or systematic. The work aimed to validate a new algorithm to generate spatial patterns for sampling soils in the context of agricultural experiments. In this sense, three functions were designed using the software R and compared with five applications (three software and two R packages). The validation was performed by replicating three spatial patterns of agricultural experiments reported in previous studies and comparing hand localization of sampling points in a sugarcane harvesting experiment with localization of points generated by the algorithm. Results show that the new algorithm can exclusively compute the localization of sampling points by experimental unit (represented by a polygon) and center these in the corresponding polygon. Other characteristics, such as computing the most common spatial point patterns and generating points along crop lines, are also founded in the other applications. Regarding in-field validation, the average distance between points generated by the algorithm and those localized manually in the field is 2.58 m. The average distance between hand located points and the closest crop row line is 0.46 m. In conclusion, the new algorithm allows locating sampling points in specific sites of the plots as the highest or lowest part of crop rows.

Keywords: furrow, precision agriculture, R software, Real Time Kinematic (RTK), soil sampling, spatial variability



Introducción

Un patrón espacial de muestreo de suelos es la forma en la que están organizados los puntos de muestreo de suelos en el espacio (Cochran, 1977). Entre los patrones espaciales más comunes se encuentran el sistemático, el completamente aleatorio, el estratificado o por unidad de área y el aleatorio sistemático (de Gruijter et al., 2006; Environmental Protection Agency [EPA], 2002; Wang et al., 2012). Estos se utilizan para localizar puntos en el terreno donde se extraen muestras de suelo que permitan analizar la variabilidad espacial y/o temporal de las propiedades del suelo (Britton, 2014; Ferguson & Hergert, 2009).

En experimentos agrícolas, la variabilidad espacial del suelo es un factor importante para entender los cambios de las variables de respuesta (Gómez & Gómez, 1984; Heil & Schmidhalter, 2017). Si el experimento involucra propiedades del suelo como variables de respuesta o covariables, la ubicación exacta de los puntos de muestreo en el terreno es fundamental (Delmelle, 2014). Por ejemplo, en experimentos donde se mide el efecto del cambio de densidad aparente del suelo al paso de la maquinaria agrícola sobre el cultivo, se deben ubicar dos puntos de muestreo (antes y después del tráfico) en sitios perpendiculares a la dirección del surco y a pocos centímetros uno de otro (Barik et al., 2014), donde efectivamente las llantas de las máquinas ejercieron presión sobre el suelo (Badalíková, 2010). Lo anterior solo se logra con una planeación milimétrica del patrón espacial de muestreo de suelo.

El procedimiento general para elaborar un patrón espacial de muestreo de suelo en un experimento agrícola es: selección del tipo de patrón espacial; selección del número de puntos de muestreo por unidad experimental y dibujo de las unidades experimentales, además de la ubicación exacta de los puntos de muestreo en cada unidad experimental en un software de dibujo asistido por computador (DAC) o de sistemas de información geográfico (SIG) (de Gruijter et al., 2006; Wang et al., 2012). El procedimiento para ubicar los puntos en campo consiste en convertir estos a un formato de archivo compatible (shapefile, texto plano, dxf, entre otros) y transferirlos a un receptor GNSS (Global Navigation Satellite System, por sus siglas en inglés) con corrección en tiempo real para localizarlos punto a punto con el receptor (Awange, 2012) y extraer la muestra de suelo. En algunos casos, el patrón espacial de muestreo se realiza para un grupo de propiedades del suelo, pero en otros, se realiza propiedad por propiedad, pues la heterogeneidad espacial de cada variable es diferente.

El procedimiento para generar un patrón espacial de muestreo implica usar recursos humanos, de tiempo y económicos (EPA, 2002). Esta es una actividad compleja que podría conducir a errores en la ubicación de los puntos (de Gruijter et al., 2006). La situación es más crítica cuando, ya diseñada la cuadrícula, se tienen que realizar cambios sobre el diseño del experimento (por una variación del número o tamaño de las unidades experimentales, aumento de las variables de respuesta o el cambio de terreno), lo que conduce a rediseñar el patrón espacial de muestreo total o parcialmente. Un nuevo experimento implica también un nuevo diseño de patrón espacial.

Son pocas las investigaciones recientes relacionadas con el tema que reportan con detalle el tipo de patrón, el número y el método de localización de muestras de suelo (Barik et al., 2014;

Driemeier et al., 2016; Qing-Huo et al., 2016). Lo anterior se explica posiblemente porque: a) el análisis estadístico tradicional de un experimento agrícola carece del contexto espacial (Montgomery, 2001), lo que, en la práctica, conduce a tomar pocas muestras de suelo sin ubicarlas espacialmente; b) hasta hace algunos años no existían herramientas para ubicar de forma precisa un punto en el terreno en tiempo real (Borgelt et al., 1996), y c) realizar un diseño de un patrón espacial de muestreo es costoso (Domburg et al., 1997).

En contraposición a estos tres argumentos se debe mencionar que: a) complementario al análisis estadístico tradicional existe el análisis geoestadístico, que permite utilizar la variabilidad intrínseca del suelo no como un factor aleatorio, sino como “la realización de variables aleatorias correlacionadas” (Bishop & Lark, 2006), además de visualizar cambios espaciales de la variable entre unidades experimentales; b) actualmente hay receptores GNSS con corrección RTK (Real Time Kinematic) con capacidad de ubicar un punto en el terreno con un error menor a 0,02 m (Rovira et al., 2010), y c) es probablemente más costoso repetir un experimento desde el inicio por errores en el muestreo que tomar una cantidad suficiente de puntos de muestreo desde el inicio.

En este contexto, la automatización del diseño de patrones espaciales de muestreo es una necesidad para aquellos investigadores que constantemente están ejecutando experimentos agrícolas, donde las propiedades del suelo son variables de respuesta o covariables que tienen un contexto espacial. No obstante, son pocas las investigaciones que han desarrollado avances tendientes a la automatización en la generación de estos patrones. Existen trabajos en temas adyacentes, como la automatización de patrones espaciales estadísticos (Brus, 2015; Castilla et al., 2014; Ripley, 1979; Wang et al., 2013), automatización en la localización, toma de muestras de suelo y medición de variables continuas en campo (Adamchuk et al., 2004; Adamchuk et al., 2005; Dalmis et al., 2016; Lobsey et al., 2010) y en la trazabilidad de las muestras (Covely, 2012). La motivación de este trabajo surgió a partir de la ausencia de herramientas para diseñar el patrón espacial del muestreo de suelos enfocadas en experimentos agrícolas. El objetivo de este trabajo fue validar un algoritmo replicando tres patrones espaciales de suelo en experimentos agrícolas reportados en investigaciones anteriores y evaluar su precisión en un experimento sobre el efecto del tráfico controlado de cosechadoras sobre las propiedades de un suelo cultivado con caña de azúcar.

Materiales y métodos

Descripción del algoritmo

Se diseñó un algoritmo en el software R versión 3.4.1 (2017-06-30) (R Core Team, 2019), consistente en tres funciones principales, más trece funciones auxiliares para distribuir patrones espaciales de muestreo de suelos de forma sistemática, aleatoria y a lo largo de surcos o líneas de cultivo. El algoritmo utiliza la librería “rgdal” (versión 1.2-8) para cargar el shapefile de líneas (que representan los surcos del cultivo) y de polígonos (que representa el lote o las parcelas experimentales). Para generar patrones aleatorios o sistemáticos, el algoritmo ejecuta el siguiente procedimiento:

- Para un muestreo aleatorio, crea n puntos, todos con coordenada de origen $(0,0)$. Si el muestreo es en cuadrícula sistemática, crea una cuadrícula con m filas y n columnas a la distancia entre filas y columnas indicadas por el usuario. Esta cuadrícula tiene como punto de origen la coordenada $(0,0)$ y su ángulo de inclinación es 0 rad. Luego ubica cada punto en una coordenada aleatoria (dentro de las coordenadas mínimas y máximas de cada polígono). El algoritmo chequea si esta coordenada se encuentra dentro del polígono. Si no, la descarta, selecciona otra y la vuelve a chequear, hasta que la coordenada efectivamente se encuentre dentro del polígono.
- Para un muestreo sistemático, el algoritmo ubica la cuadrícula en la coordenada de origen (si fue proporcionada por el usuario) y la inclina al ángulo proporcionado por el usuario. Si el usuario no proporcionó coordenada de origen, la cuadrícula se ubica en el centroide del polígono. Si hay puntos duplicados, el algoritmo selecciona uno de ellos de forma aleatoria y elimina el resto.

Para generar patrones espaciales sobre líneas de cultivo, el algoritmo realiza el siguiente procedimiento:

- Establece la distancia en la cual estará ubicado cada punto con respecto al inicio de la línea. Si el usuario proporcionó porcentajes de distancia en vez de distancia en metros, los convierte en distancia en metros, de lo contrario trabaja con las distancias en metros.
- Crea una matriz de distancia relativa, distancia absoluta y ángulo de segmentos de cada línea.
- Selecciona el segmento de la línea que contiene la distancia del punto. Calcula la nueva coordenada a partir de la coordenada inicial del segmento, la diferencia de distancia entre el punto y la distancia acumulada del segmento y el ángulo del segmento.

Variables de entrada y salida

La función para crear patrones espaciales en cuadrícula se denominó *sptpat* (Structured Point Pattern) (ecuación 1). Esta función crea una matriz de puntos a partir de un número de filas, columnas y la distancias entre filas y columnas, cuyo primer punto está ubicado en la coordenada de origen $(0,0)$. Esta matriz de puntos se puede trasladar a unas coordenadas X_0, Y_0 o se puede ubicar en el centroide de un polígono que la circunscriba. Adicionalmente, el usuario puede rotar la matriz un ángulo determinado. En esta función, es posible eliminar los puntos duplicados.

$$\text{sptpat}(\text{angle}, X_0, Y_0, \text{matrix}, \text{pol}, \text{dir}, \text{remove_duplicates}) \quad \text{Ecuación 1}$$

Donde,
angle

Número real entre -2π y 2π con ángulo de inclinación de la cuadrícula.

X0	Número real positivo con la coordenada X del punto de origen de la cuadrícula.
Y0	Número real positivo con la coordenada Y del punto de origen de la cuadrícula.
matrix	Matriz 2x2. En la primera fila contiene el número de filas y columnas (números enteros positivos). En la segunda fila contiene la distancia entre filas y columnas (números reales positivos).
pol	Nombre (sin extensión .shp) escrito entre comillas dobles (“ ”) del <i>shapefile</i> de polígonos en coordenadas planas.
dir	Dirección (entre comillas dobles “ ”) del directorio donde se encuentra el <i>shapefile</i> de polígonos.
remove_duplicates	Opción para remover duplicados (booleano TRUE/FALSE, por defecto TRUE).

La función para distribuir aleatoriamente puntos dentro de un polígono se denominó rptpat (Randomized PoinT Pattern) (ecuación 2). Esta consiste en crear un número determinado de puntos ubicados aleatoriamente dentro de un polígono dado por el usuario. Al igual que en la anterior función, en esta es posible eliminar los puntos duplicados.

rptpat(nspo, pol, dir, remove_duplicates) *Ecuación 2*

Donde,	
nspo	Entero positivo con el número de objetos espaciales (Number of SPatial Objects).
angle	Número real entre -2π y 2π con ángulo de inclinación de la cuadrícula.
pol	Nombre (sin extensión .shp) escrito entre comillas dobles (“ ”) del <i>shapefile</i> de polígonos en coordenadas planas.
dir	Dirección (entre comillas dobles “ ”) del directorio donde se encuentra el <i>shapefile</i> de polígonos.
remove_duplicates	Opción para remover duplicados (booleano TRUE/FALSE, por defecto TRUE).

La función para distribuir puntos a lo largo de surcos se denominó pclpat (Points along CropLines Pattern) (ecuación 3). Esta consiste en distribuir puntos sobre una capa de líneas de cultivos de acuerdo con un vector de distancias (o de porcentajes) proporcionado por el usuario.

pclpat(croplines, dir, w, units) *Ecuación 3*

Donde,	
croplines	Nombre (sin extensión .shp) escrito entre comillas dobles (“ ”) del <i>shapefile</i> de líneas de surcado en coordenadas planas.
dir	Dirección (entre comillas dobles “ ”) del directorio donde se encuentra el <i>shapefile</i> de líneas de surcado.
w	Vector con distancias o porcentajes en los cuales se colocará cada uno de los puntos sobre las líneas (números reales positivos).
units	Cadena de caracteres con dos opciones: “percentage” o “meters”.

La salida de cada una de las tres funciones consiste en dos vectores con las coordenadas X e Y de los puntos del patrón de muestreo.

Alcance y limitaciones

El alcance global de la herramienta está dado por el entorno en el que fue creada (software R), el cual permite ejecutar funciones como estas en la línea de comando una vez estén cargadas en el espacio de trabajo (*Workspace*). En términos de los parámetros de entrada, todos los objetos geográficos deben estar en coordenadas planas (unidad de medida metros), las unidades de longitud deben estar expresadas en metros y los ángulos en radianes. Para el caso de la función *pcflat*, la longitud, forma, punto de inicio y de finalización de cada línea de surcado (parámetro de entrada *croplines*) están definidas por un objeto geográfico de tipo *SpatialLineDataFrame*, lo cual tiene como ventaja que los puntos iniciales o finales de los surcos de un lote no tengan que ser colineales, es decir, que estén ubicados sobre la misma línea de referencia, y que los surcos puedan ser líneas rectas o curvas.

Las funciones *sptpat* y *rptpat* permiten ingresar múltiples polígonos por cada ejecución. La salida de las tres funciones no es compatible (directamente) con receptores GNSS RTK, por lo que se deben exportar los puntos resultantes (utilizando la librería “*rgdal*”) a *shapefile*. Los puntos resultantes no contienen base de datos, solo coordenadas XY.

Comparación del algoritmo con otras aplicaciones

Se realizó una búsqueda de software, aplicaciones y librerías de R con características similares, entre los cuales se encontraron los programas QGIS, Grass GIS, ArcGIS, y las librerías de R “*spatstat*” y “*splancs*”. Se instalaron las últimas versiones de estos programas y se revisó el manual de funciones de las librerías de R. Específicamente, se compararon las siguientes características: ejecución de diferentes tipos de patrones espaciales de muestreo, capacidad de generar puntos sobre líneas, capacidad de generar patrones (sistemáticos y aleatorios) por unidad de área (lote, parcela experimental, zona de manejo, etc.), inclinación de cuadrícula sistemática y generación de cuadrícula sistemática centrada en unidad de área.

Validación del algoritmo

Replicación de patrones espaciales

Se seleccionaron los patrones espaciales de muestreo de suelos reportados por Bishop y Lark (2006), Barik et al. (2014) y Driemeier et al. (2016). El patrón reportado por Bishop y Lark (2006) consistió en un doble muestreo aleatorio de 144 puntos de muestreo, cada uno en un cuadrado de 100 m de lado. Para generar este patrón se creó en QGIS un cuadrado con origen (0,0) de 100 m de lado. Posteriormente, se exportó y se ejecutó la función *sptpat* dos veces sobre el mismo polígono. El resultado de cada ejecución se convirtió en formato *shapefile*.

El patrón reportado por Driemeier et al. (2016) consistió en puntos de muestreo distribuidos en una cuadrícula de 50 x 50 m con submuestras a 5 y 10 m sobre los cuatro puntos cardinales en un campo cultivado con caña de azúcar. Para generar este patrón se localizó el área experimental en Google Earth con las coordenadas proporcionadas en el trabajo, se digitalizó y exportó en formato kml, se importó en QGIS y se exportó nuevamente en formato *shapefile*. Para generar el punto de muestreo se ejecutó la función *sptpat* con la coordenada inicial $X0=5670036$, $Y0=7632195$, ángulo $\pi/10$. Para generar los puntos de submuestreo se generaron ocho cuadrículas alrededor de la cuadrícula inicial, dos por cada punto cardinal (orientado al ángulo de inclinación de la cuadrícula inicial) a 5 y 10 m de la coordenada inicial.

El patrón reportado por Barik et al. (2014) consistió en una cuadrícula sistemática de 25 x 25 m (5 filas por 6 columnas) para evaluar el efecto de la cosecha de maíz en la compactación de suelos. Debido a que la compactación se evalúa sobre el entresurco (fondo del surco), se generó un algoritmo en R para generar surcos y entresurcos en un polígono cerrado. Se seleccionó en Google Earth un lote de la zona del estudio que cumplía con las características descritas en el trabajo (dado que este solo proporcionó las coordenadas generales de la zona), las cuales eran pendiente menor a 2 %, forma cuasi-rectangular y área de 1 ha. El lote seleccionado se digitalizó y exportó en formato kml, se importó en QGIS y se exportó en formato *shapefile*. Posteriormente, se ejecutó el algoritmo en R para crear surcos y entresurcos sobre el polígono del lote seleccionado. Después de chequear las dimensiones del lote, se ajustó el patrón a una cuadrícula de 15 m x 15 m, dado que el lote tenía una longitud de surco menor a la del trabajo. Se ejecutó la función *pclpat* para crear puntos sobre los entresurcos utilizando los siguientes parámetros: `pclpat("Intercrop_lines",dir,c(10,25,40,55,70),"meters")`. Luego se borraron los puntos sobre las líneas no distanciadas a 15 m con respecto a la línea inicial (línea 12).

Validación en campo

Se comparó el resultado del algoritmo con la ubicación manual de puntos en campo en un experimento de cosecha en caña de azúcar (12 meses de edad) con 12 unidades experimentales, ubicado en la Hacienda Potosí Suerte 001, del Ingenio Incauca (Puerto Tejada, Colombia), sobre las coordenadas $76^{\circ}25'07''W$, $03^{\circ}17'44''N$, y 963 m.s.n.m. Se ubicaron 12 puntos por parcela experimental agrupados en tres sitios, para un total de 144 puntos de muestreo, cada uno con cuatro puntos formando un cuadrado de 1 m de lado sobre la parte alta del surco de caña. Los tres grupos se localizaron sobre los surcos 6-7, 12-13 y 18-19 a 30, 60 y 90 m, respectivamente, del inicio del surco.

La ubicación manual consistió en medir con decámetro (de 10 m de longitud) desde el inicio del surco hasta la distancia del surco correspondiente y ubicar estacas de madera sobre cada uno de los cuatro puntos, las cuales luego fueron georreferenciadas con una estación RTK Hiper V (Topcon) (error promedio 0,02 m). La ubicación con el algoritmo consistió en ejecutar la función *pclpat* con el nombre del *shapefile* de líneas de surco, el directorio donde estaba almacenado el *shapefile* de líneas de surco, el vector `c(30,31,60,61,90,91)` que representa las distancias de los puntos desde el inicio de cada surco y la opción "meters", indicando que las unidades de los valores del vector estaban en metros.

Como el resultado de la ejecución del algoritmo ubica puntos sobre cada línea de cultivo, se borró manualmente los puntos diferentes a los surcos 6-7, 12-13 y 18-19. Se midió la distancia entre cada punto generado con el algoritmo y su correspondiente manual, así como la distancia más corta de cada punto ubicado manualmente a la línea de surco. Se calcularon las distancias mínimas, promedio (equivalente al error medio absoluto) y máxima entre los puntos medidos manualmente y los calculados por el algoritmo, así como la raíz cuadrada del error cuadrático medio (RECM). El procesamiento estadístico de los datos fue realizado con el software R.

Resultados y discusión

Generación de patrones espaciales

Utilizando las tres funciones se generaron los siguientes patrones a nivel de lote o de parcela experimental: muestreo en cuadrícula sistemática (figura 1a), muestreo en cuadrícula escalonada (eliminando diagonales) (figura 1b), muestreo aleatorio en cuadrícula sistemática (figura 1c), muestreo aleatorio por zona de manejo (figura 1d), muestreo en cuadrícula sistemática por parcela (figura 1e), muestreo híbrido (por zona de manejo y cuadrícula) (figura 1f), muestreo aleatorio por parcela (figura 1g) y muestreo por línea de surcado (figura 1h).

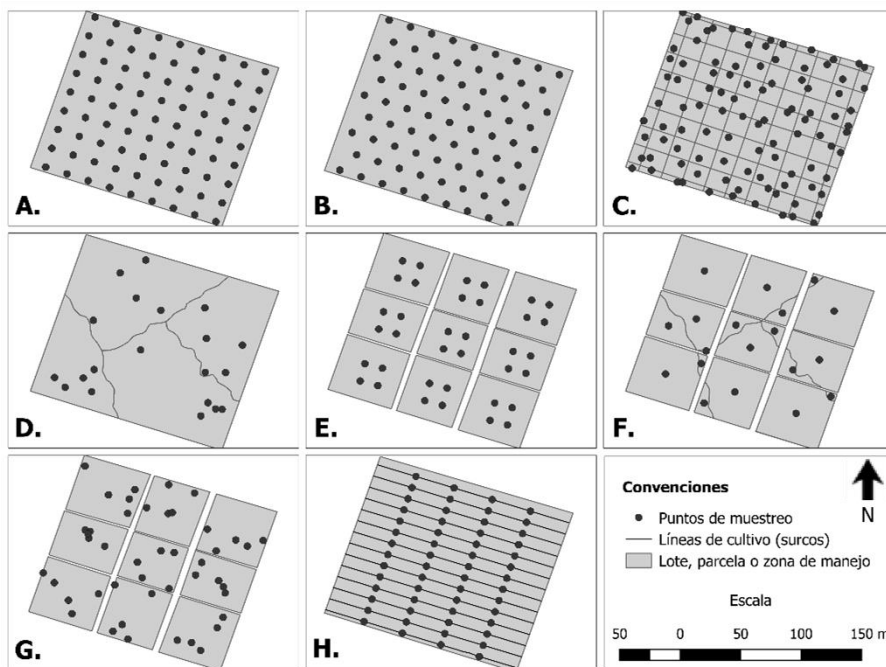


Figura 1. Representación de los patrones de muestreo obtenidos con el algoritmo. a. Aleatorio; b. Escalonado; c. Aleatorio en cuadrícula; d. Aleatorio por zona de manejo; e. En cuadrícula por parcela; f. Híbrido (por zona de manejo y parcelas); g. Aleatorio por parcela; h. Líneas de surco. Fuente: Elaboración propia

El tiempo de ejecución de las funciones para generar estos patrones fue menor a 1 s. Se observa diversidad en la distribución de los puntos en el espacio en diferentes unidades de área y a lo largo de líneas de surcos, similar a los lineamientos de muestreo para agricultura de precisión descritos por Ferguson y Hergert (2009) y a los patrones presentados por EPA (2002). La representación espacial de estos patrones a través de un algoritmo como el propuesto en este trabajo ofrece la ventaja para el investigador de que le permite planear eficientemente los muestreos de suelo evitando errores al localizar manualmente los puntos en campo, en consistencia con lo planteado por Kerry et al. (2010), y minimizar el error de interpolación espacial (Delmelle, 2014), cuando estos puntos se utilizan para generar mapas de variabilidad espacial.

Otra ventaja de generar los patrones con el algoritmo propuesto en el software R es que las coordenadas resultantes pueden ser utilizadas en un código de R para un cálculo posterior. La desventaja es que se debe tener un conocimiento básico en el manejo del software R para ejecutar las funciones.

Comparación con otras aplicaciones

En comparación con los programas QGIS, Grass GIS, ArcGIS, y las librerías de R “spatstat” y “splancs”, el algoritmo desarrollado tiene la capacidad exclusiva de realizar muestreos sistemáticos por unidad de área y centrarlos en el polígono correspondiente (figura 1e). En los otros programas y librerías también es posible generar muestreos sistemáticos, pero sin ligarlos a una unidad de área como lo hace el algoritmo desarrollado, sino a un único punto de origen o centroide (característica que el algoritmo también tiene). Esta desventaja de los programas se evidencia en los patrones de muestreo de suelo que generó Britton (2014) utilizando los programas ArcGIS y AgStudio. El resto de las características (como generación de los demás patrones y generación de puntos sobre líneas) es posible generarlas con los otros programas o librerías de R.

Con respecto al tipo de formato de salida, los programas QGIS, ArcGIS y Grass GIS generan un *shapefile* con el resultado del patrón espacial, mientras que el algoritmo y las dos librerías de R generan los patrones en formatos propios de R, como el *data.frame*. Esto es una ventaja de los programas con respecto al algoritmo y las librerías de R, siempre que el patrón sea el resultado final del proceso, lo cual no siempre sucede porque a los puntos del muestreo de suelo usualmente se les vincula la base de datos con las variables medidas en laboratorio para realizar mapas de variabilidad espacial o utilizarlas como variables de respuesta o covariables de un diseño experimental (Domburg et al., 1997). Si los puntos representan un resultado intermedio de un proceso global ejecutado en R, tener los puntos en formato *data.frame* representa una ventaja del algoritmo y las librerías de R. Igualmente, los formatos de R pueden ser exportados a formato *shapefile* o a un archivo delimitado por comas, tal como está descrito por Bivand et al. (2008).

Validación

Replicación de patrones espaciales

Para replicar el patrón espacial de muestreo presentado por Driemeier et al. (2016), se generó una cuadrícula sistemática con 202 puntos de muestreo en 50,4 ha (figura 2a). Para cada punto de muestreo se generaron ocho puntos de submuestreo, dos por cada punto cardinal inclinado sobre la cuadrícula. En total se generaron 1.616 puntos de submuestreo. La generación de este tipo de patrón espacial para muestras alteradas de suelos (con submuestreo) requirió “mover” la coordenada de origen de la cuadrícula inicial 5 y 10 m sobre cada punto cardinal.

El submuestreo, consistente en componer la muestra de suelo que representa cada punto de la cuadrícula con submuestras tomadas alrededor de este (Ferguson & Hergert, 2009), evita, por ejemplo, muestrear el suelo en puntos “calientes” (de Gruijter et al., 2006), es decir, puntos alterados significativamente por eventos naturales o antrópicos tales como una sobrefertilización, la deposición de un animal o un problema puntual de drenaje, todo lo cual ocasionaría una interpretación errónea de la variable muestreada en ese punto. La ventaja de realizar un submuestreo es que permite obtener una alta representatividad de la variable del suelo muestreada en el área de estudio cuando no hay información secundaria del suelo levantada previamente. La desventaja es el alto costo que implica su extracción.

Sin embargo, el diseño, la cantidad de submuestras y la distancia de estas al punto de muestreo principal deben ser cuidadosamente calculados, para evitar subestimar el valor de la variable distribuida espacialmente. Lo anterior se encuentra en la misma línea de lo encontrado por Rocco et al. (2016), quienes compararon la concentración de algunos metales pesados en el suelo en muestras compuestas versus la de muestras puntuales y concluyeron que en algunas de las muestras compuestas se subestima la concentración de metales pesados en relación con los valores obtenidos en las muestras puntuales.

Para replicar el patrón espacial de muestreo presentado por Bishop y Lark (2006), se generaron dos muestreos aleatorios con 144 puntos cada uno en un rectángulo de 100 x 100 m (figura 2b). La generación de este tipo de patrón se aplica en un muestreo de dos fases para observar la variación de una propiedad del suelo antes y después de que una actividad antrópica o un fenómeno natural ocurra. En el contexto de la variabilidad espacial de suelos, la principal ventaja de este patrón de muestreo es que la distancia entre los puntos no es rígida, lo cual permite capturar variabilidad a cortas distancias y evitar muestrear a lo largo de patrones sistemáticos, como el paso de un rodado o una línea de aplicación de fertilizante (Oliver & Webster, 2015), distinto a la cuadrícula sistemática reportada por Driemeier et al. (2016). La cantidad de puntos independientes del muestreo es un factor determinante para distribuir espacialmente la variable muestreada, pues de acuerdo con Oliver y Webster (2014), un número de puntos de muestreo menor a 100 podría generar un error significativamente grande en la interpolación espacial.

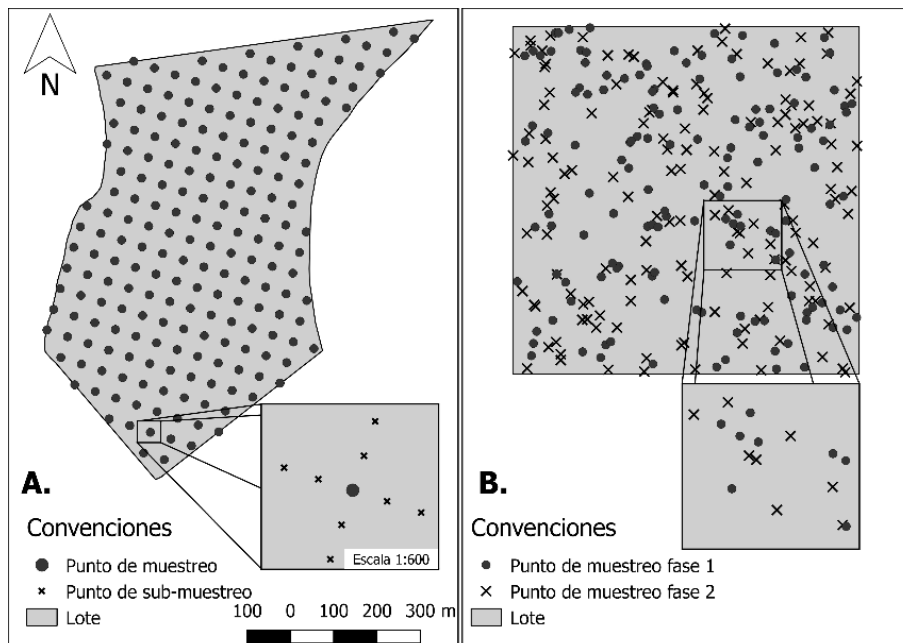


Figura 2. Replicación de patrones espaciales de muestreo de suelos presentados por Driemeier et al. (2016) (a) y por Bishop y Lark (2006) (b).

Fuente: Elaboración propia

Para replicar el patrón espacial de muestreo presentado por Barik et al. (2014), se generó una cuadrícula de 25 m x 25 m con los puntos ubicados sobre el fondo del surco (figura 3). Este tipo de patrón de muestreo se aplica cuando la evaluación de las propiedades del suelo se realiza sobre una parte específica del cultivo, en este caso el fondo del surco. Debido a que Barik et al. (2014) no reportaron cómo generaron el patrón, ni cómo localizaron los puntos en campo, se asume que la localización se realizó de forma manual (localizando los puntos en campo y luego georreferenciándolos). Una de las ventajas de generar el patrón con el algoritmo propuesto es que este se puede introducir en un receptor GNSS RTK para ubicar los puntos en campo en menor tiempo y con mayor precisión.

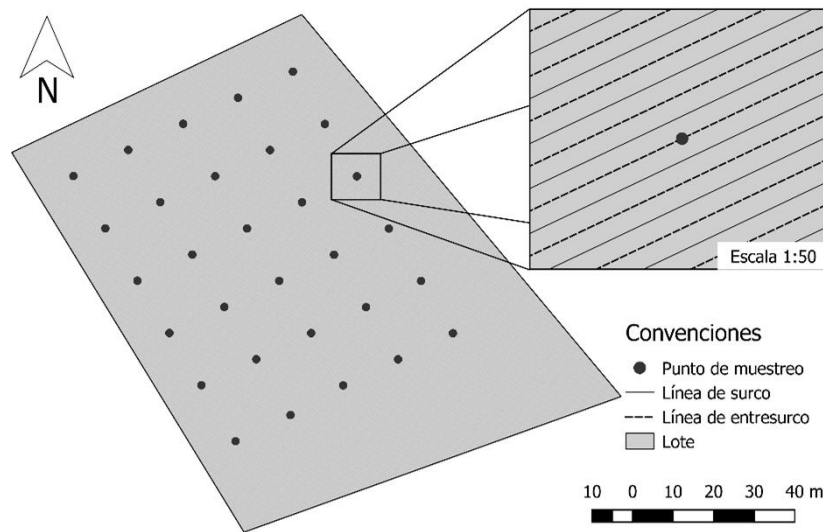


Figura 3. Replicación del patrón espacial de muestreo de suelos presentado por Barik et al. (2014).

Fuente: Elaboración propia

En la evaluación de la compactación por el tránsito de diferentes tipos de llantas de vagones de cosecha de caña de azúcar realizado por Braunack (2004), se observó que la ubicación precisa de los puntos de medición de la resistencia a la penetración, ubicados al fondo del surco, es fundamental para determinar si existen diferencias o no entre los tipos de llantas evaluadas. Este tipo de experimentos implica un reto mayor con respecto a la ubicación de los puntos de medición o muestreo, dado que es imposible dejar marcas físicas (estacas, banderillas, etc.) en campo porque la maquinaria las destruye a su paso.

Validación en campo

La comparación visual de los puntos ubicados manualmente en campo con respecto a los generados por el algoritmo permite inferir que hay diferencias entre los dos métodos (figura 4). El método de ubicación manual de puntos no es exacto debido a que la alta densidad de biomasa del cultivo de caña de azúcar y su gran altura de tallos —la cual puede llegar hasta los 3 m (Cordeiro et al., 2007)— impiden tener una referencia visual dentro del campo para ubicarse de forma exacta con respecto al mismo surco desde el inicio hasta el final de la medición.

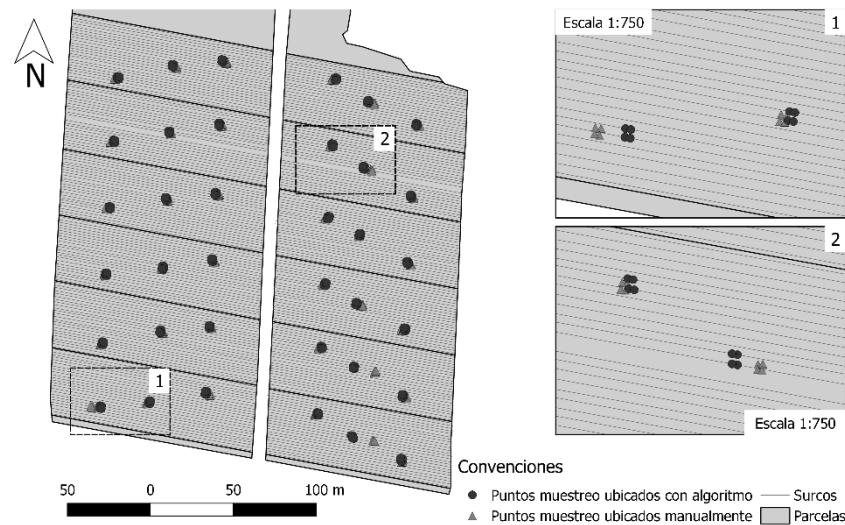


Figura 4. Distribución en campo de los puntos de muestreo ubicados con el algoritmo y los ubicados manualmente en el lote experimental.

La dificultad de ubicar con precisión puntos de muestreo dentro de lotes cultivados también es reportada por Kidd et al. (2015), quienes tuvieron que reubicar puntos de muestreo de suelo ubicados inicialmente dentro de cultivos de alto valor para evitar daños y pérdida de rendimiento. Además, las mediciones sucesivas con el decámetro desde el inicio del surco hasta el punto de muestreo generan un error instrumental mayor que la ubicación de los puntos con un receptor GNSS RTK. Es por esto por lo que algunos puntos ubicados manualmente están distanciados a más de 10 m con respecto al punto exacto generado por el algoritmo (figura 4). Dean et al. (2017) recomendaron identificar en la etapa de planeación de experimentos fuentes de error, como la que se presenta al ubicar manualmente los puntos de muestreo, las cuales podrían aumentar el error total del diseño experimental.

El resultado del análisis de distancia indica que el 62 % de los puntos ubicados manualmente se encuentran entre 0 y 2 m con respecto al punto correspondiente generado con el algoritmo, el 27 % está entre 2 y 4 m y el 6 % entre 4 y 6 m. El 5 % restante se ubicó a una distancia mayor a 10 m (figura 4). En promedio, la distancia entre los puntos ubicados manualmente con respecto a los puntos generados por el algoritmo es 2,58 m y la RECM para las coordenadas Este y Norte es 3,65 m y 0,93 m, respectivamente (tabla 1).

Tabla 1. Estadística de la distancia entre puntos ubicados manualmente y con el algoritmo y de la distancia entre los puntos manuales y el surco más cercano

Distancia (m)	Mínima	Promedio	Máxima	RECM Este	RECM Norte
Entre puntos ubicados manualmente y con el algoritmo	0,15	2,58	14,09	3,65	0,93
Entre puntos manuales a línea de surco más cercana	0,01	0,46	0,93	No aplica	No aplica

Fuente: Elaboración propia

El resultado del análisis de distancia entre cada punto ubicado manualmente y la línea de surco más cercana indica que solo uno de los 144 puntos se ubicó a una distancia menor a 0,02 m (distancia que corresponde al error de ubicación de los puntos generados por el algoritmo si se ubicaran con un receptor GNSS RTK). Solo 10 puntos se ubicaron entre 0,02 y 0,1 m, mientras los 133 restantes se ubicaron a una distancia mayor a 0,1 m (figura 5). En promedio, la distancia entre los puntos ubicados manualmente con respecto a la línea de surco más cercana es 0,46 m (tabla 1).

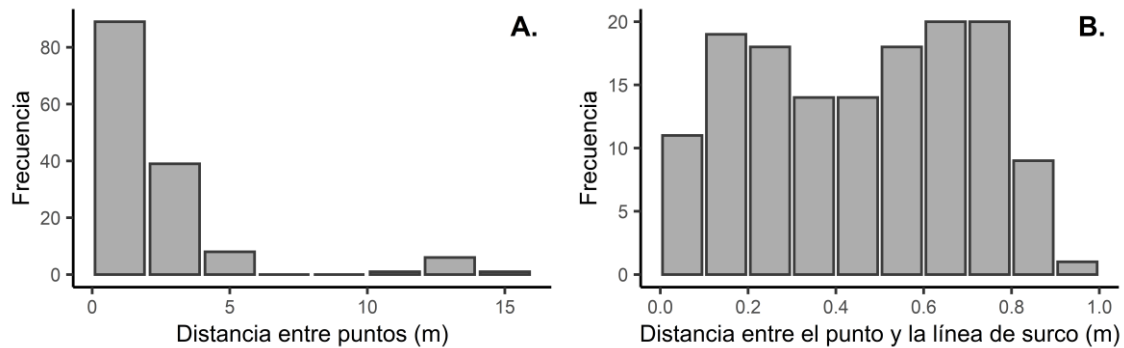


Figura 5. Histogramas de frecuencia de la distancia entre cada punto ubicado con el algoritmo y su correspondiente ubicado manualmente (a) y la distancia entre el punto ubicado manualmente y la línea de surco más cercana (b).

Teniendo en cuenta que un alto porcentaje de los puntos ubicados manualmente se localizaron a una distancia superior al ancho de la llanta de la cosechadora de caña usada en el experimento (0,36 m) (figura 5b), la ubicación manual de los puntos de muestreo de suelo impediría evaluar correctamente el efecto del tráfico de las cosechadoras autoguiadas con piloto automático sobre las propiedades físicas del suelo, objetivo central de este experimento. La ubicación de los puntos de muestreo o medición de propiedades físicas del suelo en experimentos de evaluación del tráfico de maquinaria agrícola como el validado en este estudio requieren de una planeación tan detallada que Soares et al. (2014) acuñaron el término *cama de siembra de caña* (*sugarcane seedbed region*, en inglés) para denominar la franja o región en dirección del surco delimitada por 0.40 m a la derecha e izquierda de la cepa de caña que las llantas de los tractores no deberían transitar. Finalmente, en el caso de lotes con líneas de siembra no exactamente rectas (por ejemplo, las trazadas por una sembradora sin sistema autoguía o en curvas de nivel), se recomienda ubicar manualmente los puntos de muestreo en campo y georreferenciar estos con un receptor GNSS RTK para ubicarlos posteriormente si así lo demanda el experimento.

Conclusiones

Se desarrolló y validó un nuevo algoritmo para generar patrones espaciales de muestreo de suelos en experimentos agrícolas. Este genera los patrones espaciales más comunes para muestrear el suelo, generados también por otros programas y librerías de R, pero tiene la

capacidad exclusiva de generar cuadrículas sistemáticas por unidad de área y centrar estas en el área correspondiente. Además, este logró replicar exitosamente patrones espaciales presentados en investigaciones previas.

El algoritmo permite ubicar puntos de muestreo en sitios específicos del campo, como lo son la parte alta del surco o en el entresurco, donde se muestrea el contenido de nutrientes del suelo y las propiedades físicas del suelo relacionadas con la compactación, con una precisión superior a la ubicación manual de puntos de muestreo con decámetro. Este nuevo algoritmo se constituye en una nueva herramienta para investigadores que requieran un alto nivel de precisión en la ubicación de puntos de muestreo de suelos.

Agradecimientos

Los autores agradecen al Departamento de Agricultura de Precisión del Ingenio Incauca S. A. por proporcionar y autorizar el uso de los datos para validar en campo el algoritmo.

Descargos de responsabilidad

Los autores están de acuerdo con la publicación del presente artículo y declaran que no existe ningún conflicto de interés que afecte los resultados.

Referencias

- Adamchuk, V. I., Hummel, J. W., Morgan, M. T., & Upadhyaya, S. K. (2004). On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture*, 44(1), 71-91. <https://doi.org/10.1016/j.compag.2004.03.002>
- Adamchuk, V. I., Lund, E. D., Sethuramasamyraja, B., Morgan, M. T., Dobermann, A., & Marx, D. B. (2005). Direct measurement of soil chemical properties on-the-go using ion-selective electrodes. *Computers and Electronics in Agriculture*, 48(3), 272-294. <https://doi.org/10.1016/j.compag.2005.05.001>
- Awange, J. L. (2012). *Environmental monitoring using GNSS*. Springer. <https://doi.org/10.1007/978-3-540-88256-5>
- Badalíková, B. (2010). Influence of soil tillage on soil compaction. En A. P. Dedousis, & T. Bartzanas (Eds.), *Soil Engineering* (pp. 19-30). Springer.
- Barik, K., Aksakal, E. L., Islam, K. R., Sari, S., & Angin, I. (2014). Spatial variability in soil

- compaction properties associated with field traffic operations. *Catena*, 120, 122-133. <https://doi.org/10.1016/j.catena.2014.04.013>
- Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2008). Spatial data import and export. En *Applied Spatial Data Analysis with R* (pp. 81-111). Springer. <https://doi.org/10.1007/978-0-387-78171-6>
- Bishop, T. F. A., & Lark, R. M. (2006). The geostatistical analysis of experiments at the landscape-scale. *Geoderma*, 133(1-2), 87-106. <https://doi.org/10.1016/j.geoderma.2006.03.039>
- Borgelt, S. C., Harrison, J. D., Sudduth, K. A., & Birrell, S. J. (1996). Evaluation of GPS for applications in precision agriculture. *Applied Engineering in Agriculture*, 12(6), 633-638. <https://doi.org/10.13031/2013.25692>
- Braunack, M. V. (2004). A tyre option for sugarcane haulout trucks to minimise soil compaction. *Journal of Terramechanics*, 41(4), 243-253. <https://doi.org/10.1016/j.jterra.2004.03.001>
- Britton, J. (2014). *An evaluation of soil sampling methods in support of precision agriculture in Northeastern North Carolina* (Tesis de maestría, University of Sourthen California). Digital Library. <https://digitallibrary.usc.edu/Share/6030sh0y60en8tl7vh0n8nm15sbcieh0>
- Brus, D. J. (2015). Balanced sampling: A versatile sampling approach for statistical soil surveys. *Geoderma*, 253-254, 111-121. <https://doi.org/10.1016/j.geoderma.2015.04.009>
- Castilla, G., Hernando, A., Zhang, C., Mauro, F., & McDermid, G. (2014). POLS: A versatile tool for sampling polygon GIS layers. *Computers and Geosciences*, 67, 139-149. <https://doi.org/10.1016/j.cageo.2013.10.003>
- Cochran, W. G. (1977). *Sampling techniques* (3.^a ed.). John Wiley & Sons.
- Cordeiro, G., Amouyal, O., Elliott, F., & Henry, R. (2007). Sugarcane. En C. Kole (Ed.), *Genome mapping and molecular breeding in plants: Pulses, sugar and tuber crops* (pp. 175-203). Springer.
- Covely, T. W. (2012). Patent N.º US 8,286,857 B2. <https://portal.unifiedpatents.com/patents/patent/US-8286857-B2>
- Dalmis, I. S., Akdemir, B., Turen, N., & Belliturk, K. (2016). Development of an automatic soil sampling machine. *Journal of Tekirdag Agricultural Faculty*, 13(3), 1-12. http://bhi.nku.edu.tr/basinyonetim/resim/images/edittorresimleri/522/files/01_savas_dalmis.pdf
- de Gruijter, J., Bierkens, M., Brus, D., & Knotters, M. (2006). *Sampling for natural resource monitoring*. Springer. <https://doi.org/10.1007/3-540-33161-1>
- Dean, A., Voss, D., & Draguljić, D. (2017). Planning experiments. In *Design and analysis of experiments* (2.^a ed., pp. 7-30). Springer. https://doi.org/10.1007/978-3-319-52250-0_1
- Delmelle, E. M. (2014). Spatial sampling. En M. M. Fischer, & P. Nijkamp (Eds.), *Handbook of Regional Science* (pp. 1385-1399). Springer. https://doi.org/10.1007/978-3-642-23430-9_73
- Domburg, P., de Gruijter, J. J., & Van Beek, P. (1997). Designing efficient soil survey schemes with a knowledge-based system using dynamic programming. *Geoderma*, 75(3-4), 183-201. [https://doi.org/10.1016/S0016-7061\(96\)00090-0](https://doi.org/10.1016/S0016-7061(96)00090-0)
- Driemeier, C., Ling, L. Y., Sanches, G. M., Pontes, A. O., Magalhães, P. S. G., & Ferreira, J. E. (2016). A computational environment to support research in sugarcane agriculture. *Computers and Electronics in Agriculture*, 130, 13-19. <https://doi.org/10.1016/j.compag.2016.10.002>
- Environmental Protection Agency [EPA]. (2002). *Guidance on choosing a sampling design for environmental data collection for use in developing a quality assurance project plan* [U.S. Environmental Protection Agency Quality System Series]. EPA. <https://www.epa.gov/quality/guidance-choosing-sampling-design-environmental-data-collection-use-developing-quality>

- Ferguson, R. B., & Hergert, G. W. (2009). Soil sampling for precision agriculture. *Precision Agriculture* [Universidad de Nebraska]. <https://cropwatch.unl.edu/documents/Soil%20Sampling%20for%20Precision%20Agriculture%2C%20EC154.pdf>
- Gomez, K. A., & Gomez, A. A. (1984). *Statistical procedures for agricultural research* (2.^a ed.). John Wiley & Sons. https://pdf.usaid.gov/pdf_docs/PNAAR208.pdf
- Heil, K., & Schmidhalter, U. (2017). Improved evaluation of field experiments by accounting for inherent soil variability. *European Journal of Agronomy*, 89, 1-15. <https://doi.org/10.1016/j.eja.2017.05.004>
- Kerry, R., Oliver, M. A., & Frogbrook, Z. L. (2010). Sampling in precision agriculture. En M. A. Oliver (Ed.), *Geostatistical Applications for Precision Agriculture* (pp. 35-62). Springer. <https://doi.org/10.1007/978-90-481-9133-8>
- Kidd, D., Malone, B., McBratney, A., Minasny, B., & Webb, M. (2015). Operational sampling challenges to digital soil mapping in Tasmania, Australia. *Geoderma Regional*, 4, 1-10. <https://doi.org/10.1016/j.geodrs.2014.11.002>
- Lobsey, C., Viscarra, R., & McBratney, A. (2010). An automated system for rapid in-field soil nutrient testing. En R. J. Gilkes & N. Prakongkep (Eds.), *19th World Congress of Soil Science, Soil Solutions for a Changing World* (pp. 9-12). <https://www.iuss.org/19th%20WCSS/Symposium/pdf/1.5.1.pdf>
- Montgomery, D. (2001). *Design and Analysis of Experiments* (5.^a ed.). John Wiley & Sons.
- Oliver, M. A., & Webster, R. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena*, 113, 56-69. <https://doi.org/10.1016/j.catena.2013.09.006>
- Oliver, M. A., & Webster, R. (2015). *Basic steps in geostatistics: The variogram and kriging*. Springer. <https://doi.org/10.1007/978-3-319-15865-5>
- Qing-Huo, L., Hong, L., Bao-Guo, L., Peng-Tao, G., Wei, L., & Zhao-Mu, L. (2016). Assessment of spatial uncertainty for delineating optimal soil sampling sites in rubber tree management using sequential indicator simulation. *Industrial Crops and Products*, 91, 231-237. <https://doi.org/10.1016/j.indcrop.2016.07.030>
- Soares, G., Menezes, Z., Barboza, R., Sobreira, R., & Silva, F. (2014). Effects of traffic control on the soil physical quality and the cultivation of sugarcane. *Revista Brasileira de Ciência do Solo*, 38(1), 135-146. <https://doi.org/10.1590/s0100-06832014000100013>
- R Core Team. (2019). R: *A language and environment for statistical computing*. <https://www.r-project.org/>
- Ripley, B. D. (1979). Simulating spatial patterns: Dependent samples from a multivariate density. *Journal of the Royal Statistical Society*, 28(1), 109-112.
- Rocco, C., Duro, I., Di Rosa, S., Fagnano, M., Fiorentino, N., Vetromile, A., & Adamo, P. (2016). Composite vs. discrete soil sampling in assessing soil pollution of agricultural sites affected by solid waste disposal. *Journal of Geochemical Exploration*, 170, 30-38. <https://doi.org/10.1016/j.gexplo.2016.08.004>
- Rovira Más, F., Zhang, Q., & Hansen, A. (2010). Global Navigation Systems. En *Mechatronics and intelligent systems for off-road vehicles* (pp. 43-73). Springer.
- Wang, J.-F., Jiang, C.-S., Hu, M.-G., Cao, Z.-D., Guo, Y.-S., Li, L.-F., Liu, T.-J & Meng, B. (2013). Design-based spatial sampling: Theory and implementation. *Environmental Modelling & Software*, 40, 280-288. <https://doi.org/10.1016/j.envsoft.2012.09.015>
- Wang, J.-F., Stein, A., Gao, B.-B., & Ge, Y. (2012). A review of spatial sampling. *Spatial Statistics*, 2(1), 1-14. <https://doi.org/10.1016/j.spasta.2012.08.001>