



# DESARROLLO DE UNA HERRAMIENTA GEOESTADÍSTICA PARA ELABORAR MAPAS DE VARIABILIDAD ESPACIAL DE SUELOS

Edwin Erazo Mesa<sup>1</sup>✉, Diego Sandoval López<sup>2</sup>

<sup>1</sup>. Centro de Investigación  
de la Caña de Azúcar de  
Colombia – Cenicaña.  
Estación Experimental vía  
Cali-Florida km 26

<sup>2</sup>. Ingenio Providencia S.A.

✉: eerazo@cenicana.org

**Palabras clave:**

variabilidad espacial de  
suelos, caña de azúcar,  
software R.

## RESUMEN

*El objetivo de este trabajo fue desarrollar una herramienta (rutina) para automatizar la elaboración de mapas de variabilidad espacial utilizando el software estadístico libre “R” con tres módulos interconectados. Se comparó tiempo de procesamiento y error de predicción de 102 mapas generados por la rutina y por el software ArcGis®. Para esto se utilizaron datos georreferenciados de 808 muestras de suelo sembrado con caña de azúcar distribuidos en seis cuadrículas sistemáticas de 60x60 m. de la Hacienda San Jerónimo (320 ha), ubicada en El Cerrito (Colombia). Los resultados indican que el tiempo de procesamiento de los mapas es significativamente menor utilizando la herramienta diseñada (5.40 min.) comparada con el software ArcGis (58.79 min.). Además, no hubo una diferencia significativa del error de predicción entre las dos aplicaciones para cada variable. En conclusión esta herramienta disminuye el tiempo de procesamiento de grandes volúmenes de datos georreferenciados y mantiene el error de predicción de los mapas de variabilidad espacial.*

## DEVELOPING A GEOSTATISTICAL TOOL FOR MAPPING SOIL SPATIAL VARIABILITY

**Key words:**

soil spatial variability,  
sugarcane, R software.

**SUELOS  
ECUATORIALES**  
45 (1): 36-40

ISSN 0562-5351

## ABSTRACT

*This study aimed to design a geostatistical tool that automates the preparation of spatial variability maps for soils, using freely available R software with three interconnected modules to develop a mapping routine. The runtimes and errors of prediction of 102 maps prepared using the developed routine were compared with the same maps prepared using ArcGis software. Georeferenced data sets of 808 soil samples were used. Samples were collected in sugarcane fields, distributed in six 60 x 60 m systematic grids, at the 320-ha San Jerónimo Sugarcane Plantation, located in the rural area of Cerrito, department of Valle del Cauca, Colombia. Results show mapping runtime was significantly lower with the geostatistical tool developed (5.40 min) as compared with ArcGis software (58.79 min). In addition, no significant differences were observed in error of prediction between tools. In conclusion, the geostatistical tool developed decreases runtime of large amounts of georeferenced data, while maintaining the error of prediction of spatial variability maps.*

Rec.: 02.03.2015

Acep.: 06.05.2015

## INTRODUCCIÓN

Los mapas de variabilidad espacial de suelos son representaciones gráficas de los cambios en el terreno de propiedades cuantitativas del suelo. Estos son usados por agricultores para tomar decisiones agronómicas en labores de agricultura de precisión (Sevier & Suk, 2003; Oliver, 2010). En caña de azúcar se usan principalmente para generar el plan de fertilización en tasa variada (Mosquera, 2012), que consiste en generar un mapa con diferentes dosis de fertilizante dentro del campo de acuerdo a la variabilidad de las propiedades del suelo y la productividad del cultivo.

Para elaborarlos se utiliza un software con capacidad de análisis geoestadístico como ArcGis, Surfer, SAGA, Matlab, SAS o R (Goovaerts, 2010). En cada software se debe seleccionar paso a paso los parámetros del análisis exploratorio de datos, del semivariograma experimental, el modelo de semivariograma teórico de mejor ajuste, el método de interpolación adecuado y los parámetros de evaluación para la validación cruzada (Hengl, 2004; Robinson & Metternicht, 2006). Se deben seleccionar al menos diez parámetros para elaborar un mapa de variabilidad espacial. Esto es una tarea compleja para usuarios poco experimentados.

Hay softwares comerciales diseñados para procesar grandes volúmenes de datos georreferenciados por lotes (batch processing) como SST Software (SST Software, 2015), pero su alto costo de adquisición, requerimiento de área mínima a trabajar (2000 ha), acceso y uso compartido de los datos, así como la necesidad de conexión a internet para procesar los datos en la nube, entre otros, limitan su uso a pequeñas agroindustrias y agricultores. Hay herramientas automáticas para generar mapas de variabilidad espacial en Web como INTAMAP (Interoperability and Automated Mapping) (Pebesma et al, 2009) pero no procesa por lotes. En resumen, no hay herramientas de uso directo con una relación costo/beneficio favorable para generar mapas de variabilidad espacial a partir de datos georreferenciados.

El objetivo de este trabajo fue desarrollar una herramienta para automatizar la elaboración de mapas de variabilidad espacial. Esto representa la primera aproximación a solucionar el problema de procesar grandes volúmenes de datos georreferenciados de variables ambientales generados en ingenios y por cultivadores del sector agroindustrial de la caña de azúcar de Colombia.

## MATERIALES Y MÉTODOS

### Recolección de datos.

El muestreo de suelo se realizó entre agosto y diciembre de 2012 en la hacienda San Jerónimo (320 ha), ubicada en El Cerrito (Colombia) sobre 76°20'45" W y 3°37'40" N. Para recolectar las 808 muestras se trazó una cuadrícula sistemática (Britton, 2014) no continua de 60 x 60 m. Las muestras de suelo se extrajeron ajustando la metodología propuesta por Ferguson & Hergert (2009). Cada muestra se compuso de cinco submuestras tomadas en el punto central y a 10 m. de éste en los cuatro puntos cardinales, a una profundidad de 0.30 m. Dada la no continuidad espacial de la cuadrícula sistemática en el área de estudio ésta se dividió en seis sectores cada uno con más de cien puntos de muestreo (Figura 1).

Para cada muestra de suelo se determinó pH, conductividad eléctrica (Cea), materia orgánica (MO), fósforo (P), calcio (Ca), magnesio (Mg), Potasio (K), Sodio (Na), capacidad de intercambio catiónico (CIC), Boro (B), Hierro (Fe), manganeso (Mn), cobre (Cu), zinc (Zn) y porcentajes de arena (A), limo (L) y arcilla (Ar) utilizando protocolos de laboratorio para análisis de suelos cultivados con caña de azúcar (Cenicña, 2009).

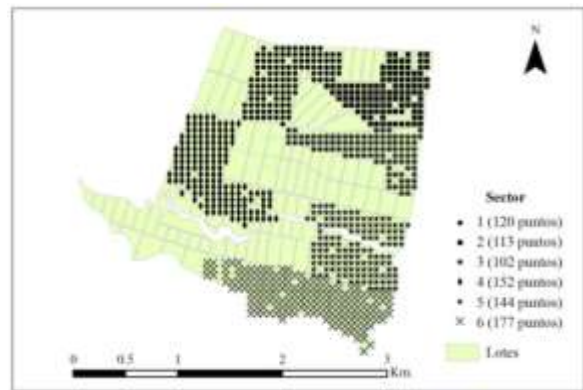


Figura 1. Localización de puntos de muestreo por sectores en la Hacienda San Jerónimo.

### Herramienta geoestadística.

La herramienta se diseñó basada en el marco metodológico propuesto por Hengl (2009) que establece, a través de un árbol de decisión, que cualquier variable ambiental distribuida espacialmente a mapear cuyo modelo físico se desconozca y no se correlacione con otras covariables ambientales puede ser mapeada de dos formas: con un interpolador mecánico como *Spline* o *Inverse Distance Weighting (IDW)* si no presenta autocorrelación espacial o con Kriging Ordinario si la presenta. A partir de esto se escribió una rutina

computacional con tres módulos en lenguaje R utilizando el software R v. 3.1.0 (R Core Team, 2014).

El primer módulo (análisis exploratorio de datos) identifica y descarta los valores atípicos de cada variable, tendencias espaciales y puntos duplicados, entre otros. Calcula también el número de observaciones totales y de datos faltantes, además de estadísticos descriptivos como el promedio, desviación estándar, coeficiente de variación, varianza, mínimo, primer cuartil, mediana, segundo cuartil, máximo, coeficiente de asimetría, tipo de transformación y curtosis.

El segundo (análisis semivariográfico) calcula el semivariograma experimental de cada variable y lo ajusta a diez modelos teóricos: Esférico, Exponencial, Gaussiano, Circular, Matern, Lineal, Nugget, Pentaesférico, Periódico y Logarítmico. Luego selecciona automáticamente el de mejor ajuste basado en el menor valor numérico de *Akaike Information Criterion* (AIC), un criterio que mide el ajuste relativo de un modelo estadístico a un conjunto de datos (McBratney y Webster, 1986).

El tercero (interpolación y validación cruzada) define la cuadrícula de interpolación a partir de un origen de coordenadas, tamaño y número de celdas. Luego interpola con IDW si el modelo escogido es “Nugget” o con Kriging Ordinario si es diferente. Finalmente calcula la validación cruzada a través de los parámetros raíz del error cuadrático medio (RMSE), el error medio (ME), la raíz del error cuadrático medio normalizado (RMSEr) y coeficiente de determinación entre los valores medidos y estimados ( $r^2$ ) (Hengl, 2009).

**Evaluación de la herramienta.**

Para evaluar el tiempo de procesamiento de 102 mapas (17 propiedades del suelo x 6 sectores) se

estableció un diseño experimental completamente al azar con dos tratamientos (herramienta R y extensión *Geostatistical Analyst* de ArcGis v. 10.2.), cada uno con tres repeticiones. Un usuario con experiencia en el manejo de software SIG ejecutó cada módulo de la herramienta tres veces. Posteriormente el mismo ejecutó tres veces la extensión de ArcGis 102 veces. Las pruebas se realizaron en horas de la mañana.

La herramienta geoestadística generó una base de datos de los errores de predicción por sector y por variable. En ArcGis se almacenó el shapefile de la validación cruzada de cada mapa. Para analizar los parámetros de la validación cruzada (RMSE, ME, RMSEr y  $r^2$ ) se estableció un diseño experimental completamente al azar con los mismos tratamientos de la anterior evaluación pero tomando cada sector como una repetición. Todos los mapas se generaron a un tamaño de pixel de 20x20 m. en formato raster (ASCII). Los datos de entrada para cada herramienta fue la base de datos de suelos georreferenciada (shapefile de puntos para ArcGis y un archivo de texto delimitado por comas (.csv) para R) y el dato de salida es un mapa en formato raster con sus parámetros de validación cruzada.

**RESULTADOS Y DISCUSIÓN**

Cada software produjo 102 mapas de variabilidad espacial interpolados con Kriging Ordinario debido a que los semivariogramas experimentales presentaron autocorrelación espacial. La herramienta geoestadística tiende a “suavizar” más la superficie interpolada con respecto al *Geostatistical Analyst* de ArcGis. En el mapa de pH elaborado con ambas herramientas (Figura 2) se observan similitudes en la variabilidad espacial de la zona norte y oeste de los mapas (sectores 1 a 4) y diferencias en la zona sur (sectores 5 y 6).

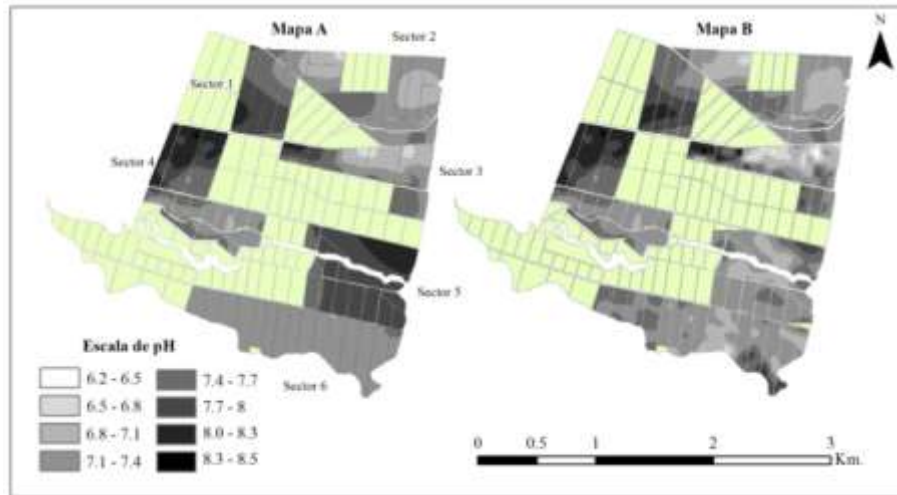


Figura 2. Mapa elaborado con la herramienta geoestadística (Mapa A) y con ArcGis (Mapa B).

El tiempo de procesamiento de los mapas es significativamente menor utilizando la herramienta diseñada (promedio 5.40 min.) comparada con el software ArcGis (promedio 58.79 min.) (Tabla 1). La prueba de Tukey (95% de probabilidad) para comparación de múltiples medias entre tratamientos indica que hay una diferencia altamente significativa tanto en el tiempo de procesamiento total de los 102 mapas como en el tiempo de procesamiento por mapa en ambos tratamientos ( $p_{adj.}=0.000$  en ambas pruebas).

Se observa que para la mayoría de las variables, ArcGis genera los mapas con menor error (RMSE), mayor correlación ( $r^2$ ) y mayor grado de explicación de la variabilidad a través de la interpolación (RMSEr) con respecto a la herramienta geoestadística, sin embargo la prueba de Tukey para comparación de múltiples medias entre tratamientos indica que no hay diferencias significativas entre ambos tratamientos (softwares) para los errores evaluados (Tabla 2).

**Tabla 1.** Tiempos de elaboración de los mapas utilizando las dos herramientas

Tiempo (min.)	Herramienta geoestadística - R			Geostatistical Analyst - ArcGis		
	Rep. 1	Rep. 2	Rep. 3	Rep. 1	Rep. 2	Rep. 3
Procesamiento total	5.38	5.40	5.42	63.48	58.77	54.12
Procesamiento promedio por mapa	0.04	0.05	0.05	0.62	0.58	0.53

**Tabla 2.** Errores de predicción promedio de los mapas utilizando las dos herramientas

Variable	Tratamiento 1 (R)				Tratamiento 2 (ArcGis)				Prueba de Tukey ( $p_{adj.}$ )			
	RMSE <sup>1</sup>	ME	RMSEr	$r^2$	RMSE	ME	RMSEr	$r^2$	RMSE	ME	RMSEr	$r^2$
Arcilla (%)	3.32	0.00	61.34	0.62	3.23	0.00	59.33	0.64	0.73**	0.89**	0.70**	0.69**
Arena (%)	6.14	0.02	59.33	0.64	5.91	0.03	57.19	0.67	0.67**	0.93**	0.62**	0.58**
B (ppm)	0.13	0.00	90.81	0.34	0.13	0.00	90.42	0.18	0.97**	0.56**	0.90**	0.26**
Ca (cmol/Kg)	3.19	-0.01	75.34	0.43	3.09	0.01	73.08	0.46	0.87**	0.42**	0.48**	0.51**
Cea (dS/m)	0.22	0.00	145.12	0.30	0.11	0.00	86.83	0.24	0.35**	0.39**	0.31**	0.72**
CIC(cmol/Kg)	5.80	-0.01	76.14	0.41	5.59	0.00	74.20	0.44	0.89**	0.82**	0.73**	0.76**
Cu (ppm)	0.92	0.00	79.15	0.36	0.93	-0.01	79.58	0.36	0.87**	0.29**	0.95**	0.96**
Fe (ppm)	8.96	-0.02	76.90	0.41	8.67	-0.08	74.12	0.45	0.67**	0.30**	0.34**	0.39**
K(cmol/Kg)	0.28	0.00	81.10	0.34	0.27	0.00	79.71	0.35	0.92**	0.68**	0.84**	0.90**
Limo (%)	5.00	-0.03	68.87	0.52	5.08	-0.03	68.23	0.52	0.88**	0.80**	0.93**	1.00**
Mg(cmol/Kg)	1.22	-0.01	67.84	0.53	1.18	-0.01	66.00	0.55	0.84**	0.69**	0.76**	0.79**
Mn (ppm)	3.24	0.00	91.19	0.33	3.23	-0.01	90.81	0.18	1.00**	0.48**	0.94**	0.34**
Mo(%)	0.40	0.00	80.99	0.34	0.39	0.00	80.62	0.35	0.98**	0.49**	0.88**	0.90**
Na (cmol/Kg)	0.13	0.00	89.94	0.69	0.12	0.00	81.49	0.31	0.72**	0.77**	0.44**	0.08**
P (ppm)	41.56	-0.19	76.16	0.40	41.40	-0.48	75.70	0.41	0.98**	0.35**	0.95**	0.96**
pH	0.25	0.00	72.67	0.46	0.24	0.00	69.87	0.50	0.77**	0.28**	0.68**	0.67**
Zn (ppm)	0.51	0.00	80.55	0.33	0.51	0.00	79.36	0.35	0.86**	0.88**	0.89**	0.90**

\*\* No hay diferencias significativas entre tratamientos

<sup>1</sup>RMSE y ME están expresados en unidades de la variable, mientras RMSEr está expresado en porcentaje.

Adicional al tiempo de procesamiento, otras características diferencian las dos herramientas. Por ejemplo, la herramienta geoestadística selecciona automáticamente el modelo teórico de semivariograma de mejor ajuste entre diez modelos. Geostatistical Analyst solamente optimiza el modelo seleccionado por el usuario. La herramienta geoestadística procesa datos por lote, es decir para generar  $n \times p$  mapas ( $n$  variables por  $p$  sectores), la herramienta se debe ejecutar una vez mientras que se debe ejecutar  $n \times p$  veces la extensión Geostatistical Analyst. Una desventaja de la herramienta geoestadística con respecto a Geostatistical Analyst es que esta no tiene una interface gráfica como la de ArcGis.

## CONCLUSIONES

- La herramienta geoestadística diseñada disminuye el tiempo de procesamiento de grandes volúmenes de datos georreferenciados y mantiene el error de predicción de los mapas de variabilidad espacial comparada con la *Geostatistical Analyst* de ArcGis.
- La herramienta geoestadística selecciona automáticamente el modelo teórico de semivariograma de mejor ajuste, procesa datos por lote, una ventaja con respecto a otras herramientas geoestadísticas. Una desventaja de la herramienta geoestadística es que no tiene una interface gráfica.
- Esta herramienta permite adicionar otras características (como anisotropía, correlación con otras covariables ambientales) para disminuir el error de predicción. Además se puede utilizar para interpolar otras variables como la productividad del cultivo de caña de azúcar.

## BIBLIOGRAFIA

- BRITTON J. (2014). An Evaluation of Soil Sampling Methods in Support of Precision Agriculture in Northeastern North Carolina. (Master of Science Thesis, University of Southern California, California, United States of America). Obtenido de <http://digitallibrary.usc.edu/cdm/ref/collection/p15799coll13/id/377297>
- CENICANA (2009). Determinación de los cationes intercambiables: Ca, Mg, K y Na en suelos por espectroscopía de absorción atómica. 7 p.
- FERGUSON, R. B., & HERGERT, G. W. (10 de Febrero de 2015). University of Nebraska Lincoln. Obtenido de Soil Sampling for Precision Agriculture: <http://www.ianrpubs.unl.edu/live/ec154/build/ec154.pdf>
- GOOVAERTS P. (2010). Geostatistical Software. En: FISCHER & GETIS (EDS.) Handbook of Applied Spatial Analysis (pp. 125-133). Vienna: Springer.
- HENGL, T., HEUVELINK, G. B., & STEI, A. (2004). A generic framework for spatial prediction of soil variables. *Geoderma*, 120, 75–93.
- HENGL T. (2009). A Practical Guide to Geostatistical Mapping. 293 p.
- MCBRATNEY A.B., WEBSTER R. (1986). Choosing functions for semivariograms of soil properties and fitting them to sampling estimates. *Journal of Soil Science* (37): 617-639 p.
- MOSQUERA C. (2012). Fertilización en tasa variada. *Revista Tecnicaña* (28): 36 p.
- OLIVER, M. A. (2010). Geostatistical Applications for Precision Agriculture. Reading: Springer.
- PEBESMA, E., CORNFORD, D., DUBOIS, G., HEUVELINK, G., HRISTOPULOS, D., PILZ, J., . . . SKOEN, J. (2011). INTAMAP: The design and implementation of an interoperable automated interpolation web service. *Computers & Geosciences* (37), 343–352.
- R CORE TEAM (12 de Febrero de 2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Obtenido de: <http://www.R-project.org/>.
- ROBINSON T., METTERNICHT G (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture* (50), 97–108 p.
- SEVIER B., SUK W. (2003). Adoption Trends and Attitudes Towards Precision Agriculture in Florida Citrus: Preliminary Results from a Citrus Producer Survey. En: ASAE Annual International Meeting Paper Number: 031100 (27- 30 July 2003: Las Vegas).
- SST SOFTWARE. (10 de Febrero de 2015). Farm Rite Product Service. Obtenido de SST Software: <http://www.sstsoftware.com/>