

# Desarrollo de un software para mapeo de variabilidad espacial en agricultura y ambiente

Paccioretti, P., Córdoba, M. y Balzarini, M.

DOI: 10.31047/1668.298x.v37.n1.27863

## RESUMEN

Diversos datos agronómicos y ambientales son recolectados con sistemas de posición geográfica posibilitando la construcción de mapas que describen la variación de las propiedades en el espacio. La variabilidad espacial uni y multivariada ha sido utilizada para comprender variación de observaciones dentro de campos de cultivo y también entre sitios de un territorio extenso, permitiendo la delimitación de áreas o zonas homogéneas a distintas escalas. El uso de métodos de estadística espacial, tanto a escala regional como a escala fina, demanda programas de computación especializados. FastMapping es una aplicación web interactiva que fue desarrollada para procesar datos espaciales. Permite preprocesar archivos de datos georreferenciados (depurar, estandarizar y re-escalar), facilita la interpolación espacial y la generación de mapas de variabilidad espacial de manera cuasiautomática. En este trabajo se comparó la zonificación multivariada de lotes agrícolas en función de propiedades de suelo producida por FastMapping con la obtenida desde un software comercial difundido para agricultura de precisión. Las facilidades de FastMapping para importar, limpiar, analizar y visualizar datos espaciales, en un único ambiente de computación, proveen un entorno fácil de usar para la implementación efectiva de métodos de estadística espacial.

**Palabras clave:** agricultura de precisión; datos espaciales; depuración; zonas de manejo.

Paccioretti, P., Córdoba, M. and Balzarini, M. 2020. Software development for spatial variability mapping in agriculture and environment. Agriscientia 37: 75-84

## SUMMARY

Diverse agronomic and environmental data are collected with geographical position systems enabling the construction of maps that describe the variation of the studied properties over the space. Uni and multivariate spatial variability has been used to understand observation variation within crop fields and between sites in a large territory, allowing the delimitation of homogeneous areas at different scales. The use of spatial statistics methods, both regionally and on a small scale, demands specialized software programs. FastMapping

is an interactive web application developed to process spatial data. It allows users to preprocess georeferenced data (clean, standardize and re-scale), it facilitates spatial interpolation and the generation of spatial variability maps quasi-automatically. This work compares the multivariate zoning of agricultural fields according to soil properties produced by FastMapping and that obtained from commercial software used for precision agriculture. FastMapping allows to import, clean, analyze and visualize spatial data, in a single environment, provide an easy-to-use framework for the effective implementation of spatial statistics methods.

**Keywords:** precision agriculture; spatial data; depuration; management zone.

*Paccioretti, P. (ORCID: 0000-0002-1267-6361): Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba, Córdoba, Argentina. Unidad de Fitopatología y Modelización Agrícola (UFYMA), Instituto Nacional de Tecnología Agropecuaria (INTA) - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Córdoba, M. (ORCID: 0000-0002-4761-9623) y Balzarini, M. (ORCID: 0000-0002-4858-4637): Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba, Córdoba, Argentina. Unidad de Fitopatología y Modelización Agrícola (UFYMA), INTA-CONICET. Correspondencia a: pablopaccioretti@agro.unc.edu.ar*

## INTRODUCCIÓN

El análisis de datos georreferenciados provee la oportunidad de modelar la variabilidad espacialmente estructurada tanto a escala de lote como a escala regional. La interpolación espacial conocida como *kriging* está basada en el ajuste de funciones de variogramas y es usada para predecir valores de una variable analizada en sitios no muestreados de un dominio espacial. Los valores predichos sobre una malla de predicción, pueden luego ser usados para generar mapas temáticos que son prácticamente continuos en el espacio (Betzek *et al.*, 2019). Numerosos modelos teóricos de variogramas (Cressie y Wikle, 2011) pueden ser ajustados desde los datos muestreados para realizar interpolaciones espaciales. La selección de un modelo en particular se puede realizar a partir de medidas que evalúan el desempeño de las predicciones espaciales (Córdoba, Paccioretti, Giannini Kurina, Bruno y Balzarini, 2019). Ajustar y evaluar cada uno de los distintos modelos de variograma puede ser complejo. La automatización del proceso para la selección del modelo espacial y la predicción necesaria para el mapeo ayuda a la implementación de métodos eficientes para el análisis de datos espaciales.

El mapeo univariado y multivariado de datos georreferenciados es útil para delimitar zonas

homogéneas, tanto a escala fina (Córdoba, Bruno, Costa, Peralta y Balzarini, 2016) como a escala regional (Giannini Kurina, Hang, Córdoba, Negro y Balzarini, 2018). Al explorar la variabilidad especial dentro de un lote cultivado, por ejemplo, los agricultores pueden identificar los factores principales que impactan en el rendimiento. Por lo tanto, un objetivo común en la agricultura basada en datos (agricultura de precisión) es la clasificación no supervisada de sitios dentro del lote para identificar áreas similares dentro del lote con respecto a los factores limitantes para el desarrollo del cultivo (Moral, Terrón y Marques da Silva, 2010). Estas áreas homogéneas (zonas de manejo) constituyen la base de la gestión agrícola sitio-específica.

Los datos provenientes de monitores de rendimiento son frecuentemente usados para analizar la variabilidad intralote y realizar zonificaciones. Sin embargo, otras capas de datos (mediciones sobre distintos sitios del lote) pueden estar disponibles; por ejemplo, mediciones de electroconductividad aparente, profundidad de suelo y propiedades edáficas. En estos contextos, de múltiples capas de datos, la zonificación multivariada es recomendable (Taylor, McBratney y Whelan, 2007). Para este fin, existe especial interés en métodos de *clúster* que consideren la correlación espacial presente entre

las observaciones de una misma variable tomada en distintos sitios del terreno (autocorrelación espacial) y la correlación entre variables (Córdoba, Bruno, Costa y Balzarini, 2013). Sin embargo, los métodos de conglomeración o *clustering* de sitios que contemplan la espacialidad de los datos, no se encuentran disponibles en la mayoría de los softwares usados para la delimitación de zonas homogéneas en sentido multivariado.

Otros problemas asociados al análisis de datos espaciales tienen que ver con el manejo de las bases de datos previo al análisis. Usualmente provienen de sistemas de información geográfica (SIG) o se encuentran ligados a sistemas de posición geográfica (GPS) que almacenan datos en formatos diversos. En casos donde los datos se colectan automáticamente (como con tecnologías de agricultura de precisión), las bases de datos pueden contener numerosos errores (Sudduth y Drummond, 2007). En este trabajo se describe una nueva aplicación de software llamada FastMapping, la cual ha sido desarrollada para automatizar, en una interfase simple y fácil de usar, un protocolo completo de análisis estadísticos para datos espaciales que incluye la automatización del preprocesado de la base de datos, la interpolación geoestadística para la construcción y visualización de mapas de variabilidad espacial y la clasificación de sitios en zonas homogéneas a partir de un algoritmo multivariado con restricción espacial. Los objetivos de este trabajo fueron analizar la variabilidad intralote del rendimiento de trigo y covariables de sitio, y comparar la zonificación multivariada producida por FastMapping con la obtenida a partir de software comercial de amplia difusión en agricultura de precisión.

## MATERIALES Y MÉTODOS

### Software

El software FastMapping fue desarrollado utilizando paquetes especializados del lenguaje de programación R (R Core Team, 2019), los cuales posibilitan la aplicación de protocolos de depuración, interpolación y clasificación uní y multivariada en el contexto de datos correlacionados espacialmente (Córdoba *et al.*, 2013; Vega, Córdoba, Castro-Franco y Balzarini, 2019). La interfaz gráfica de usuario está generada principalmente con funciones del paquete *shiny* (Chang, Cheng, Allaire, Xie y McPherson, 2019). La depuración espacial se realiza con funciones del paquete *spdep* (Bivand y Wong, 2018). El ajuste de semivariogramas y la interpolación espacial

usa los paquetes *automap* (Hiemstra, Pebesma, Twenhöfel y Heuvelink, 2009), *fields* (Nychka, Furrer, Paige y Sain, 2017), *geoR* (Ribeiro y Diggle, 2018) y *raster* (Hijmans, 2019). La clasificación es realizada con los paquetes *adespatial* (Dray, Saïd y Débias, 2008) y *e1071* (Meyer, Dimitriadou, Hornik, Weingessel y Leisch, 2019). FastMapping está alojado en un servidor de la Universidad Nacional de Córdoba (UNC). El código de programación se encuentra en <https://github.com/PPaccioretti/FastMapping>. Para comenzar una sesión de FastMapping, es necesario contar con conexión a Internet y se debe ingresar a <https://fastmapping.psi.unc.edu.ar>.

### Datos de ilustración

Se utilizó para procesar con FastMapping, una base de datos de un lote comercial (60 ha) de trigo (*Triticum aestivum* L.) bajo agricultura continua, localizado en el sureste de la provincia de Buenos Aires. Los datos incluían medidas georreferenciadas de rendimiento de trigo en una campaña. Además, se midió valores de electroconductividad aparente (ECa) en dos profundidades, 0-30cm y 0-90cm, elevación y profundidad de suelo. Los valores de rendimiento fueron obtenidos con monitores de rendimiento. Las medidas de ECa fueron tomadas usando una sonda Veris 3100 (VERIS *technologies*, Salina, KS, USA). La medición de la elevación del terreno se realizó utilizando un DGPS Trimble R3 (Trimble *Navegation Limited*, Sunnyvale, CA, USA). La profundidad del suelo fue medida usando un penetrómetro hidráulico.

### Análisis estadístico

#### *Preprocesamiento de datos*

Previo al procesamiento, se usó FastMapping para transformar el sistema de coordenadas geográficas a cartesianas Universal Transversal de Mercator (UTM). Todas las variables fueron depuradas y re-escaladas a una grilla regular de 30x30m utilizando FastMapping. En la depuración de datos, se removieron datos relacionados a los efectos bordes. Además, se eliminaron valores que se encontraban por fuera del patrón general de los datos. Para esto, el conjunto de datos se limita dentro de un rango de variación razonable, donde los valores mínimos y máximos se obtienen a partir del conocimiento previo de la distribución. Luego, se calcula la media y la desviación estándar (DE), eliminando valores que se encuentren por fuera del intervalo media  $\pm 3$  DE, dado que teóricamente

aproximadamente el 90 % de los datos de una variable aleatoria se encuentra dentro de este rango (*outlier* global). Por último, se eliminaron aquellos valores que, si bien su observación es posible para la distribución de la variable, difieren significativamente de su vecindario (*outlier* espacial). Para detectar estos valores, se utiliza el índice de autocorrelación espacial local de Moran (IM) (Anselin, 1995) y el gráfico de dispersión de Moran (Anselin, 1996). Si al menos uno de estos criterios detecta la observación como rara, ésta es eliminada automáticamente. Para calcular la autocorrelación espacial con el índice de Moran, es necesario generar una red de vecindarios, en este se consideraron sitios vecinos a aquellos que se encontraban dentro de una distancia de 35 metros.

#### Mapeo de la variabilidad espacial

##### Mapeo univariado

Para el mapeo de la variabilidad espacial, FastMapping ajusta 11 modelos teóricos de semivariogramas: *Exponential*, *Spherical*, *Gaussian*, *Matern*, *Matern Stein's*, *Circular*, *Linear*, *PowerWave*, *Pentaspheical* y *Hole* (Hiemstra *et al.*, 2009). Para la selección del modelo de mejor ajuste utiliza un proceso de validación cruzada (Kohavi, 1995) que evalúa la capacidad predictiva del modelo de correlación espacial seleccionado. Aquel con el que se obtiene menor error cuadrático medio de predicción es automáticamente seleccionado para ser usado en la interpolación espacial.

FastMapping utiliza distintos métodos de *kriging* (Webster y Oliver, 2007) para realizar interpolaciones espaciales y poder predecir los valores de la variable en sitios no muestreados. Este método proporciona el mejor predictor lineal para el valor de la variable en un sitio, suministrando además un error de predicción conocido como varianza *kriging*, que depende del modelo de variograma ajustado y de las localizaciones de los datos originales. La varianza *kriging* brinda la posibilidad de analizar la calidad de las predicciones obtenidas por interpolación. Para evitar el uso de información proveniente de muestras redundantes, el método *kriging* pondera de forma distinta muestras que están cerca entre sí respecto a muestras que están más distantes. Los parámetros del variograma ajustado determinan las ponderaciones de los datos que serán usados en la interpolación.

##### Mapeo de variabilidad multivariada

Sobre la base de datos depurada y re-escalada se pueden aplicar diferentes versiones del análisis de clúster no jerárquico *fuzzyk-means* (Bezdek, Coray, Gunderson y Watson, 1981). FastMapping permite implementar *fuzzy k-means* sobre las mediciones originalmente medidas, las obtenidas por interpolación durante el re-escalado o sobre las componentes principales espaciales del archivo en análisis (Córdoba *et al.*, 2013). Para realizar el Análisis de Componentes Principales espacial (Dray, Saïd y Débias, 2008), se requiere primero calcular la matriz de ponderación espacial en forma similar a la realizada para el cálculo del índice de Moran. Para determinar el número óptimo de clústeres, se calculan los coeficientes de partición (FPI, del inglés *fuzziness performance index*) y de entropía de la clasificación (NCE, del inglés *normalized classification entropy*) (Bezdek *et al.*, 1981) junto con el índice Xie-Beni (Xie y Beni, 1991). Además, FastMapping combina estos índices, produciendo un índice resumen, para obtener una única métrica para determinar el número de clúster óptimo (Galarza, Mastaglia, Albornoz y Martínez, 2013). Para aplicar el algoritmo *fuzzy k-means*, es necesario definir el parámetro conocido como exponente difuso. En el contexto de generación de zonas de manejo en agricultura de precisión, usualmente se utiliza un valor de 1,3 (Odeh, McBratney y Chittleborough, 1992). El algoritmo de clúster necesita la especificación de una métrica de distancia para evaluar la similitud/diferencia entre los objetos a agrupar. La métrica de distancia, Euclídea, Diagonal y de Mahalanobis son frecuentemente usadas en delimitación de zonas de manejo (Fridgen, Kitchen y Sudduth, 2004). FastMapping provee de manera automática la comparación de medias entre zonas para cada una de las variables involucradas en el análisis multivariado. La comparación de medias se realiza con varianzas calculadas después del ajuste por correlación espacial según la siguiente expresión

$$|Y_{ZonaX} - Y_{ZonaW}| \geq (\hat{\sigma}_{krig} \times 1,96) \times 2$$

donde  $Y_{ZonaX}$  es la media de la Zona X,  $Y_{ZonaW}$  es la media de la Zona W y  $\hat{\sigma}_{krig}$  es la mediana de la desviación estándar *kriging*; el valor 1,96 es el cuantil 0,975 de la distribución normal estándar que se supone siguen las diferencias de medias y el valor 2 deriva de la condición bilateral de la hipótesis sobre igualdad de medias de la variable de interés entre zonas (Taylor *et al.*, 2007).

Implementación del mapeo de variabilidad espacial en FastMapping

Al cargar el archivo de datos crudos (*Wheat raw data*, disponible en FastMapping) FastMapping transforma el sistema de coordenadas geográficas a un sistema de coordenadas cartesianas UTM, luego de especificar el hemisferio (sur) y el área (20). El usuario puede cargar una base de datos donde se especifican los vértices sobre el cual se realizará la interpolación. Cuando se selecciona una columna como *Target variable* las pestañas *Depuration*, *Prediction*, *Results*, *Cluster* y *Report* se habilitan. En la pestaña *Depuration* las opciones *Global Outliers*, *Spatial Outliers* y *Remove borders* pueden dejarse con sus valores por defecto y de esta manera, previo a realizar la interpolación, el software elimina los valores menores o iguales a cero, observaciones que se encuentran dentro de los 20 metros del borde y valores que se encuentran por fuera del intervalo de la media  $\pm 3$  DE. Para eliminar *outlier* espaciales, se consideran vecinos aquellos sitios que se encuentran hasta una distancia determinada (20 metros). En la pestaña *Prediction*, la opción correspondiente a los modelos teóricos a ajustar y si se selecciona *Automatic*, FastMapping ajustará uno a uno, 11 modelos teóricos de variograma. Para la interpolación espacial, si se selecciona la opción *Ordinary* se realizará una predicción *kriging* puntual. Para esto se definen la cantidad de vecinos de cada punto a predecir (de 7 a 25). Por defecto, la interpolación se realiza sobre una grilla de 30x30m (opción *Grid dimensions*). Estos procedimientos se repiten para cada variable de la base de datos para conformar el archivo que se debe ingresar para realizar la zonificación multivariada. El archivo denominado *Zoning a field from yield and soil properties*, disponible en FastMapping constituye un ejemplo del formato de datos necesario implementar una zonificación multivariada.

Evaluación de FastMapping

El software MZA (del inglés *Management Zone Analyst*) ha sido desarrollado para utilizar la información del lote y delimitar clases o potenciales zonas de manejo y se encuentra ampliamente difundido como software comercial para la zonificación de lotes en agricultura de precisión (Fridgen *et al.*, 2004). En este trabajo, se comparó el desempeño de FastMapping y MZA a través del análisis de la base de datos de referencia, con ambos softwares. Para el análisis de clúster en MZA se utilizó como medida de similitud la distancia Mahalanobis, la cual se recomienda cuando se utilizan variables correlacionadas y con varianzas heterogéneas como *input* del

análisis (Fridgen *et al.*, 2004). Otras opciones de configuración utilizadas fueron: número máximo de iteraciones=300 y criterio de convergencia=0,0001. El exponente difuso se fijó en el valor convencional de 1,30 (Odeh *et al.*, 1992).

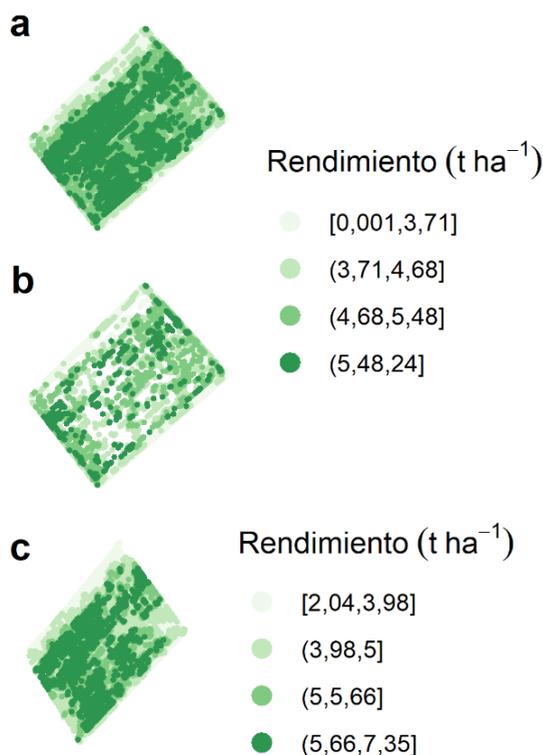
La zonificación se realizó con los datos previamente depurados y reescalados en FastMapping ya que MZA no contiene facilidades para el preprocesamiento de datos. Las visualizaciones de los resultados producidos por FastMapping se comparan con visualizaciones realizadas con gráficos generados en R para representar los clústeres de sitios producidos por MZA ya que los resultados de la zonificación de MZA se entregan en formato de un archivo de texto.

## RESULTADOS Y DISCUSIÓN

### Mapeo univariado

Previo a la depuración, los valores de rendimiento en grano de trigo se encontraban entre 0,001 t ha<sup>-1</sup> y 24 t ha<sup>-1</sup>, con media de 4,62 t ha<sup>-1</sup> y DE 2,07 t ha<sup>-1</sup>. Luego de la depuración de la base de datos, el valor mínimo de rendimiento fue de 2,04 t ha<sup>-1</sup> y el máximo de 7,34 t ha<sup>-1</sup> con media de 4,82 t ha<sup>-1</sup> y DE 1,52 t ha<sup>-1</sup>. Aun cuando la presencia de valores extremos de rendimiento no impactó significativamente en la estimación de la media, la depuración de estos valores puede cambiar la varianza de la variable en estudio y sesgar la estimación de los modelos de variabilidad espacial que serán usados para el mapeo, por ello es recomendable la eliminación de valores raros previo a la aplicación de métodos de estadística espacial. Luego de procesar numerosos mapas de rendimiento, se observa que el proceso de depuración mejora las predicciones espaciales y reduce los efectos negativos de los errores de lectura en la distribución espacial del rendimiento (Vega *et al.*, 2019). En la base de datos de trigo, sin depuración, el error de predicción fue del 20,2 % respecto a la media del rendimiento de trigo; luego de la depuración ese error disminuyó a 7,7 %. Del total de observaciones (n = 9810), el 14,2 % fueron eliminadas por encontrarse a una distancia menor a 20 metros de los bordes. En agricultura de precisión las observaciones muy cercanas a los límites del lote suelen contener errores asociados a los manejos de cabecera y giros de las maquinarias, razón por la cual se recomienda su eliminación (Sudduth y Drummond, 2007). El 1,82 % de los datos fueron considerados *outliers* globales y el 14,8 % se identificaron como

*outliers* espaciales. Luego de interpretar estas observaciones raras con el gráfico de dispersión de Moran, FastMapping eliminó automáticamente el 4,7 % de los datos. En la pestaña *Plot Condition Data* de FastMapping, pueden visualizarse el conjunto de las observaciones que se usarán en la interpolación y la posición de aquellas eliminadas por distintos motivos. En total, para el lote de trigo bajo análisis, se eliminó el 35,5 % de los sitios debido a la presencia de valores raros (Figura 1).



**Figura 1.** Gráficos mostrando sitios y su valor; a) datos crudos; b) datos eliminados por distintos motivos durante la depuración, c) observaciones no eliminadas.

Para el ajuste del semivariograma experimental, el modelo teórico que menor error de predicción generó fue el propuesto por Matern con la parametrización de Stein (Stein, 1999). Las estimaciones de los parámetros del semivariograma fueron Nugget = 0, Sill Parcial = 1,1, Rango = 68,4 y  $\kappa = 0,4$ . Estos resultados sugieren que los datos están correlacionados hasta una distancia de 68,4 m y que la variabilidad observada se encuentra espacialmente estructurada. Los valores predichos y la varianza *kriging* pueden descargarse desde la pestaña *Predicted*. Los datos extraídos durante la depuración y su condición de extracción también pueden ser descargados.

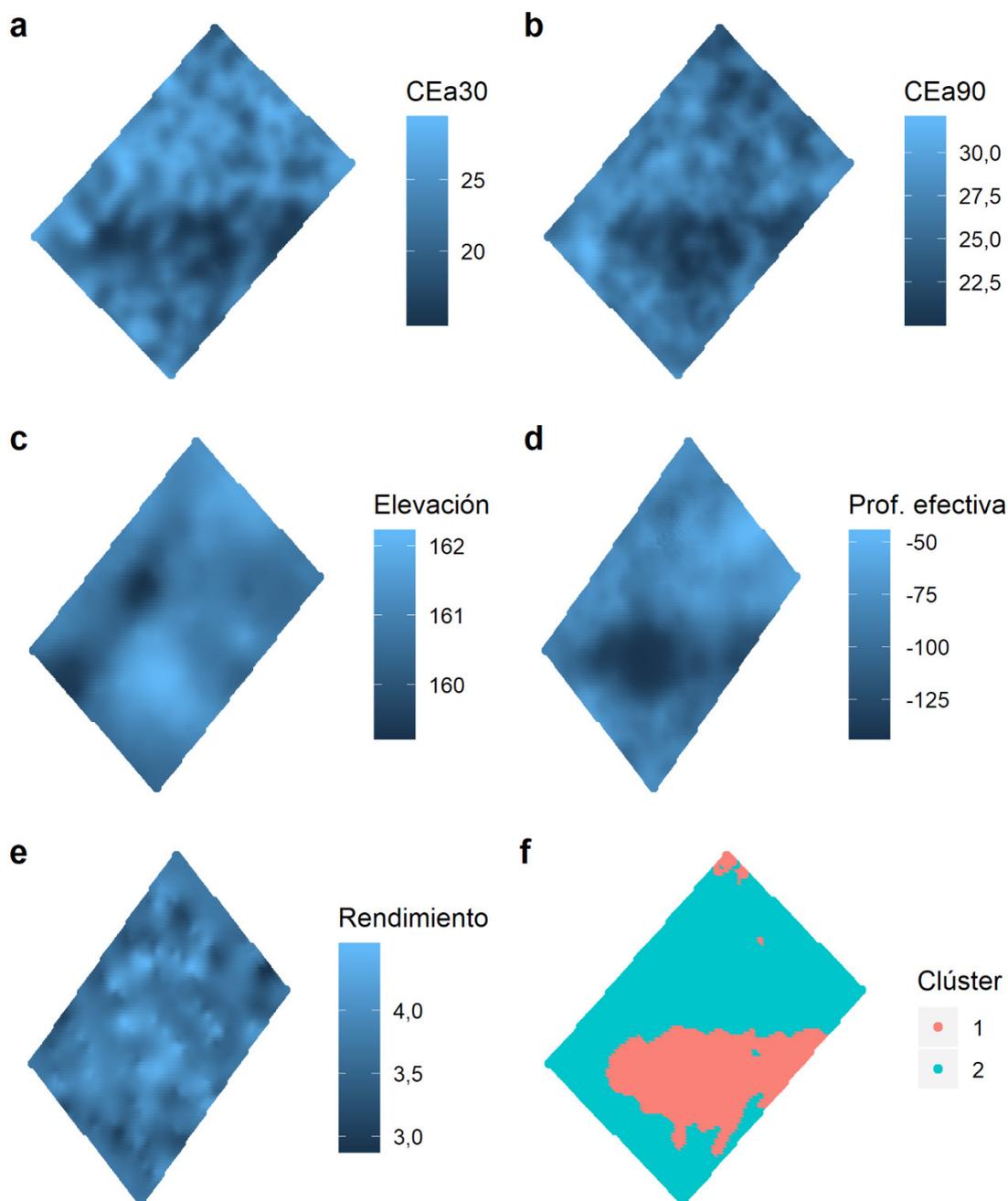
## Mapeo multivariado

Cuando se selecciona más de una variable para el análisis, FastMapping habilita la pestaña *Multivariate* posibilitando la clasificación multivariada mediante el método KM-sPC (Córdoba *et al.*, 2013). Por defecto, el software realiza la estandarización de las variables y almacena las componentes espaciales que explican hasta el 70 % de la variabilidad total. Para el lote de trigo, analizado no solo con el rendimiento sino también con las covariables de sitio descriptas en la sección Datos, el índice resumen sugirió dos conglomerados de sitios o zonas. Para la interpretación de la zonificación obtenida, se mapeó la variabilidad espacial de cada una de las covariables de sitio y del rendimiento en grano. Los resultados de la zonificación también se presentan como un mapa de sitios geoposicionados donde estos son coloreados de acuerdo con el clúster al que fueron asignados.

La variabilidad espacial de la electroconductividad aparente, en ambas profundidades (Figura 2 a y 2 b) y la variabilidad de la profundidad efectiva del suelo (Figura 2 d), ponen de manifiesto la existencia de un área con menores valores de electroconductividad aparente y mayor profundidad de suelo. Esta depresión ha sido determinante en la conformación de las zonas homogéneas (Figura 2 f), donde se observan las dos zonas delimitadas por el algoritmo de clúster KM-sPC. La variabilidad espacial de rendimiento de trigo para la campaña analizada mostró, por el contrario, menos estructuración espacial. FastMapping reportó diferencias estadísticamente significativas entre medias de zonas para todas las variables medidas a excepción del rendimiento en grano ( $\alpha = 0,05$ ).

## Comparación de FastMapping con MZA

Los procedimientos realizados con MZA requirieron más tiempo computacional que con FastMapping. Para visualizar los clústeres generados con MZA, los resultados en formato de texto debieron exportarse a otro software, mientras que en FastMapping pueden visualizarse dentro del mismo ambiente de trabajo. El software MZA reemplaza las celdas vacías por ceros, sin advertir este cambio al usuario, mientras que FastMapping, realiza el análisis sin considerar las celdas vacías. El remplazo por el valor cero, principalmente en situaciones con muchos datos faltantes, puede ocasionar problemas en las estimaciones de variabilidad espacial. FastMapping permite importar datos de diversos formatos mientras



**Figura 2.** Mapeo de la variabilidad del lote en estudio para cinco variables medidas: a) conductividad eléctrica aparente 0-30 cm; b) conductividad eléctrica aparente 0-90 cm; c) elevación; d) profundidad efectiva del suelo; e) rendimiento de trigo; f) mapeo de la variabilidad multivariada contemplando todas las características anteriores.

que MZA se restringe a datos de texto con las columnas delimitadas por coma. El análisis de los datos con FastMapping pudo realizarse sin necesidad de cambiar el ambiente de trabajo

dado que la depuración, la interpolación espacial y la clasificación multivariada están integrados en el software. La velocidad de cómputo de FastMapping dependerá de la memoria disponible

en el servidor, pero usualmente lleva algunos minutos correr un análisis completo para obtener el mapa deseado.

La zonificación producida por FastMapping fue menos fragmentada (Figura 3) y las diferencias de medias para variables de suelo entre las dos zonas delimitadas, constituyen un resultado de referencia que permite comparar los dos softwares. Las zonas definidas por FastMapping mostraron mayor diferencia de medias para todas las variables respecto a las zonas producidas con MZA (Tabla 1). Al menos, en este campo se observó una mejora de la zonificación. El algoritmo KM-sPC que implementa FastMapping para la zonificación, ha mostrado que puede producir zonas más contiguas cuando la correlación espacial entre las variables es alta (Córdoba *et al.*, 2013). Sin embargo, en varias aplicaciones de *fuzzy k-means*, que es el algoritmo implementado en MZA, se observó el inconveniente de una alta fragmentación de las zonas (Frogbrook y Oliver, 2007; Ping y Dobermann, 2003).

## CONCLUSIÓN

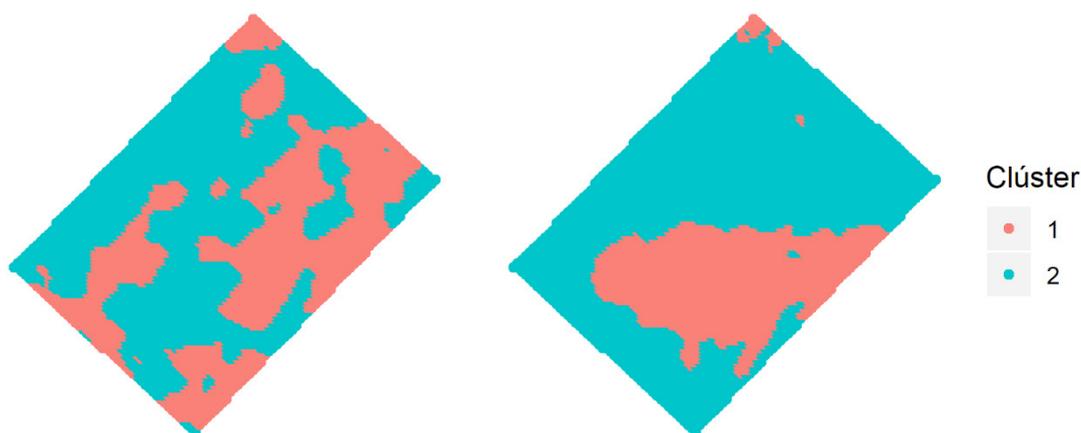
La innovación metodológica de FastMapping es la integración de una etapa de pre-procesamiento de datos orientada a eliminar anomalías en datos espaciales, la construcción y validación de mapas de variabilidad espacial y la delimitación y validación de zonas homogéneas, tanto a escala de lote como a escala regional. Aplicado en datos de agricultura

**Tabla 1.** Comparación y diferencia de medias de variables medidas entre zonas delimitadas por *Management Zone Analyst* (MZA) y FastMapping.

Software	Variable	Clúster 1	Clúster 2	Diferencia
MZA	CEa30	23,64 a	20,90 b	2,74
	CEa90	26,37 a	24,70 a	1,66
	Elevación	160,96 a	160,79 a	0,18
	Profundidad efectiva	-92,48 a	-91,78 a	0,69
	Rendimiento trigo	3,78 a	3,71 a	0,07
FastMapping	CEa30	23,97 a	18,89 b	5,08
	CEa90	26,57 a	23,47 b	3,10
	Elevación	160,75 a	161,21 b	0,47
	Profundidad efectiva	-81,49 a	-117,59 b	36,11
	Rendimiento trigo	3,71 a	3,85 a	0,15

Electroconductividad aparente a 0-30 cm (ECa30) y 0-90 cm de profundidad (ECa90), elevación (m.s.n.m.) y profundidad de suelo efectiva (cm). Para cada variable medias con letras distintas indican diferencias significativas ( $p < 0,05$ ).

de precisión FastMapping mejoró la zonificación lograda respecto a métodos convencionales de clasificación, delimitando zonas con mayores diferencias en variables de suelo y topografía. La disponibilidad de aplicaciones basadas en R, con una interfaz amigable, permite mejorar las bases metodológicas y computacionales del análisis de datos espaciales.



**Figura 3.** Mapeo de la variabilidad multivariada del lote en estudio obtenidos con *Management Zone Analyst* (MZA) (izquierda) y FastMapping (derecha).

## AGRADECIMIENTOS

Las investigaciones metodológicas que sustentan el protocolo de análisis implementado en FastMapping y su desarrollo son subsidiadas por el proyecto PID 2018 del MinCyT de la Provincia de Córdoba y por el proyecto PROIINDIT 2019 de la SeCyT de la Facultad de Ciencias Agropecuarias, UNC, Córdoba, Argentina.

## BIBLIOGRAFÍA

- Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. En Manfred Fischer, Henk J. Scholten y David Unwin (Eds.), *Spatial Analytical Perspectives on GIS Vol. 4* (p. 121-138). London, United Kingdom: Taylor & Francis.
- Betzek, N. M., Souza, E. G. de, Bazzi, C. L., Schenatto, K., Gavioli, A. y Magalhães, P. S. G. (2019). Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. *Computers and Electronics in Agriculture*, 157, 49–62. <https://doi.org/10.1016/j.compag.2018.12.004>
- Bezdek, J. C., Coray, C., Gunderson, R. y Watson, J. (1981). Detection and characterization of cluster substructure I. Linear structure: Fuzzy c-lines. *SIAM Journal on Applied Mathematics*, 40(2), 339–357.
- Bivand, R. S. y Wong, D. W. S. (2018). Comparing implementations of global and local indicators of spatial association. *TEST*, 27, 716–748. <https://doi.org/10.1007/s11749-018-0599-x>
- Chang, W., Cheng, J., Allaire, J. J., Xie, Y. y McPherson, J. shiny: Web Application Framework for R (version 1.4.0.2) [Software de cómputo]. Recuperado de: <https://cran.r-project.org/package=shiny>
- Córdoba, M., Bruno, C., Costa, J. L. y Balzarini, M. (2013). Subfield management class delineation using cluster analysis from spatial principal components of soil variables. *Computers and Electronics in Agriculture*, 97, 6–14. <https://doi.org/10.1016/j.compag.2013.05.009>
- Córdoba, M. A., Bruno, C. I., Costa, J. L., Peralta, N. R. y Balzarini, M. G. (2016). Protocol for multivariate homogeneous zone delineation in precision agriculture. *Biosystems Engineering*, 143, 95–107. <https://doi.org/10.1016/j.biosystemseng.2015.12.008>
- Córdoba, M., Paccioletti, P., Giannini Kurina, F., Bruno, C. y Balzarini, M. (2019). Guía para el análisis de datos espaciales. Aplicaciones en agricultura. Córdoba, Argentina: Brujas.
- Cressie, N., y Wikle, C. K. (2011). *Statistics for spatio-temporal data*. New York, USA: John Wiley & Sons.
- Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G... y Wagner, H. H. adespatial: Multivariate Multiscale Spatial Analysis (versión 0.3-8) [Software de cómputo]. Recuperado de: <https://cran.r-project.org/package=adespatial>
- Dray, S., Saïd, S. y Débias, F. (2008). Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. *Journal of Vegetation Science*, 19 (1), 45–56. <https://doi.org/10.3170/2007-8-18312>
- Fridgen, J. J., Kitchen, N. R. y Sudduth, K. A. (2004). Management zone analyst (MZA): *Agronomy Journal*, 96 (1), 100–108. <https://www.soils.org/publications/aj/abstracts/96/1/100>
- Frogbrook, Z. L. y Oliver, M. A. (2007). Identifying management zones in agricultural fields using spatially constrained classification of soil and ancillary data. *Soil Use and Management*, 23 (1), 40–51. <https://doi.org/10.1111/j.1475-2743.2006.00065.x>
- Galarza, R., Mastaglia, M. N., Albornoz, E. M. y Martínez, C. (2013). Identificación automática de zonas de manejo en lotes productivos agrícolas. V Congreso Argentino de Agroinformática (CAI) y 42da. Jornadas Argentinas de Informática. Córdoba, Argentina. <http://sinc.unl.edu.ar/sinc-publications/2013/GMAM13>
- Giannini Kurina, F., Hang, S., Córdoba, M. A., Negro, G. J. y Balzarini, M. G. (2018). Enhancing edaphoclimatic zoning by adding multivariate spatial statistics to regional data. *Geoderma*, 310, (170-177). <https://doi.org/10.1016/j.geoderma.2017.09.011>
- Hiemstra, P. H., Pebesma, E. J., Twenhöfel, C. J. W. y Heuvelink, G. B. M. (2009). Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Computers & Geosciences*, 35(8), 1711–1721. <https://doi.org/10.1016/j.cageo.2008.10.011>
- Hijmans, R. J. raster: Geographic Data Analysis and Modeling (versión 3.1-5) [Software de cómputo]. Recuperado de: <https://cran.r-project.org/package=raster>
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. En XXIV International Joint Conference on Artificial Intelligence, (2), 1137–1145. San Francisco, CA, United States: Morgan Kaufmann Publishers Inc.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. y Leisch, F. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien (versión 1.7-3) [Software de cómputo]. Recuperado de: <https://cran.r-project.org/package=e1071>
- Moral, F. J., Terrón, J. M. y Marques da Silva, J. R. (2010). Delineation of management zones using mobile measurements of soil apparent electrical conductivity

- and multivariate geostatistical techniques. *Soil and Tillage Research*, 106 (2), 335–343. <https://doi.org/10.1016/j.still.2009.12.002>
- Nychka, D., Furrer, R., Paige, J. y Sain, S. *fields: Tools for spatial data (version 10.3)* [Software de cómputo]. Boulder, Colorado, U.S.: University Corporation for Atmospheric Research.
- Odeh, I. O. A., McBratney, A. B. y Chittleborough, D. J. (1992). Soil pattern recognition with fuzzy-c-means: application to classification and soil-landform interrelationships. *Soil Science Society of America Journal*, 56 (2), 505–516. <https://doi.org/10.2136/sssaj1992.03615995005600020027x>
- Ping, J. L. y Dobermann, A. (2003). Creating Spatially Contiguous Yield Classes for Site-Specific Management. *Agronomy Journal*, 95 (5), 1121-1131. <https://doi.org/10.2134/agronj2003.1121>
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing (versión 3.6.3)* [Software de cómputo]. Viena, Austria. Recuperado de: <https://www.r-project.org/>
- Ribeiro, P. J. y Diggle, P. J. (2018). *geoR: Analysis of Geostatistical Data. (version 1.8-1)* [Software de cómputo] Recuperado de: <https://cran.r-project.org/package=geoR>
- Stein, M. L. (1999). *Interpolation of Spatial Data. Some Theory for Kriging*. New York, United States: Springer. <https://doi.org/10.1007/978-1-4612-1494-6>
- Sudduth, K. A. y Drummond, S. T. (2007). Yield Editor: Software for Removing Errors from Crop Yield Maps. *Agronomy Journal*, 99 (6), 1471–1482. <https://doi.org/10.2134/agronj2006.0326>
- Taylor, J. A., McBratney, A. B. y Whelan, B. M. (2007). Establishing Management Classes for Broadacre Agricultural Production. *Agronomy Journal*, 99(5), 1366-1376. <https://doi.org/10.2134/agronj2007.0070>
- Vega, A., Córdoba, M., Castro-Franco, M. y Balzarini, M. (2019). Protocol for automating error removal from yield maps. *Precision Agriculture*, 19(102), 1–15.
- Webster, R. y Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*(2da ed.). New York, USA: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470517277>
- Xie, X. L. y Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(8), 841–847. <https://doi.org/10.1109/34.85677>