

Delimitación de zonas de manejo sitio-específicas

Córdoba Mariano¹, Bruno Cecilia¹, Costas José Luis² y Balzarini Mónica¹

¹*Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba-CONICET, Córdoba, Argentina.* ²*INTA, Balcarce, Argentina*

Resumen

El manejo sitio-específico promueve la identificación y el manejo de zonas dentro del lote que presentan características homogéneas en cuanto a factores que determinan el rendimiento de los cultivos. Sin embargo, la determinación de zonas de manejo es difícil debido a la compleja combinación de estos factores y sus interacciones. Para capturar la variabilidad del rendimiento se miden atributos topográficos y/o propiedades físico-químicas del suelo. Estas variables se obtienen en múltiples sitios dentro del lote. Un aspecto importante para la delimitación de zonas de manejo es el análisis de la información recolectada de cada sitio. Dada la naturaleza multidimensional del problema (varias variables para un mismo sitio en múltiples sitios) las técnicas de análisis multivariado revisten alta utilidad. En este trabajo se ilustra la aplicación de dos técnicas multivariadas, el análisis de cluster utilizando el software Management Zone Analyst y MULTISPATI-PCA del software libre R, para la identificación de zonas de manejo en un lote en producción.

Abstract

Site-specific management allows the identification of homogeneous zones within field regarding factors that affect crop yield. However, due to the complex combination of these factors, determining management zones is a difficult task. Topographic attributes and/or soil properties are usually used to capture yield variability, obtaining several variables for each site in the field. An important aspect for the delimitation of management zones is the analysis of data gathered at each site. Given the multidimensional nature of the problem, multivariate analyses are very useful tools. In this work we illustrate the application of two multivariate techniques, cluster analysis, using Management Zone Analyst software and MULTISPATI-PCA implemented in the free software R, to identify management zones in a production field.

Palabras Clave

Análisis de componentes principales, análisis de conglomerados no jerárquico, autocorrelación espacial, sitio-específico.

Introducción

Las características del suelo y del cultivo varían en el espacio (distancia y profundidad) y en el tiempo. La agricultura de precisión es un conjunto de técnicas orientado a optimizar el uso de los insumos agrícolas en función de la cuantificación de la variabilidad espacial y temporal de la producción (Bongiovanni *et al.*, 2007). Uno de los enfoques más difundidos para el manejo de la variabilidad espacial es el manejo sitio-específico, entendiéndose por tal a la delimitación de subregiones que expresan una combinación relativamente homogénea de factores de rendimiento dentro de los lotes, tales como la textura, la topografía y el nivel de nutrientes (Moral *et al.*, 2010). El manejo sitio-específico permite mejorar la eficiencia productiva y optimizar el uso de insumos en términos agronómicos, económicos y ambientales. Sin embargo, definir zonas de manejo (ZM) suele ser complejo debido a las relaciones subyacentes entre los factores que podrían afectar el rendimiento del cultivo, como factores edáficos, meteorológicos y antropogénicos (Corwin *et al.*, 2010). Estos factores interactúan de forma dinámica, provocando que las decisiones de cómo determinar las ZM sean dificultosas. Para la delimitación de ambientes homogéneos es necesario tener en cuenta tres aspectos (Fridgen *et al.* 2004): (a) la información inicial (variables o factores utilizados como base para la delimitación de las zonas en el lote) (b)

los procedimientos para procesar e integrar la información inicial con el objetivo de definir las unidades o zonas de manejo y (c) el número de zonas de manejo en el que el lote será dividido. Diferentes enfoques básicos se han utilizado para delimitar ZM, implementando desde el uso de cartas de suelo, mapas geológicos (Kühn *et al.*, 2009), la utilización de técnicas de interpolación geoestadística para estimar la distribución espacial de las propiedades del suelo (Corwin *et al.*, 2003) hasta el análisis de datos digitales de elevación recolectados a través de sistemas de posicionamiento global (Kitchen *et al.*, 2005). Debido a que el rendimiento del cultivo es el mejor indicador de la productividad del suelo en las distintas subregiones del lote, los mapas de rendimiento son utilizados para la delineación de ZM (Jaynes *et al.*, 2005). Sin embargo, existen objeciones al uso exclusivo del rendimiento para la delimitación de ZM debido a su variabilidad temporal y a la interacción que pueda existir con el tipo de cultivo o prácticas de manejo usadas por el productor. En los últimos años ha tenido un gran crecimiento la utilización de la CE_a para inferir sobre la distribución espacial de algunas propiedades de los suelos y para delimitar zonas de manejo sitio-específica en agricultura (Corwin *et al.*, 2010; Moral *et al.*, 2010; Kitchen *et al.*, 2005). La delimitación de zonas en función de los atributos topográficos y/o propiedades físico-químicas del suelo (CE_a y elevación) frecuentemente capta la variabilidad del rendimiento debido a que se relacionan con el contenido de agua útil del suelo y por lo tanto con el potencial de producción de los cultivos (Corwin *et al.*, 2003).

Un tipo de algoritmo computacional difundido específicamente para la delimitación de ambientes, es el análisis de cluster que consiste en la identificación de los grupos, en sentido del parecido multivariado de las observaciones que los conforman. En análisis de cluster los grupos o conglomerados se arman de manera tal que las diferencias dentro de cada grupo son minimizadas y se maximizan las diferencias entre grupos. Así el sitio del lote, sobre el cual se han registrado varias variables es asignado a una clase o cluster en particular (Johnson *et al.*, 2003; Ortega y Santibáñez, 2007). Un software basado en el análisis de cluster especialmente desarrollado para la delimitación de ZM denominado MZA (por sus siglas en inglés Management Zone Analyst, Fridgen *et al.*, 2004) utiliza el algoritmo no supervisado fuzzy k-means. MZA provee los resultados de la clasificación de las ZM para una serie de números de cluster o ZM simultáneamente, lo cual permite al usuario evaluar la cantidad de ZM óptimas a través de dos índices de performance: 1) el índice de desempeño del grado de ambigüedad (Fuzziness Performance Index, FPI) y 2) entropía de clasificación normalizada (Normalized Classification Entropy, NCE). Los análisis multivariados usualmente utilizados en agricultura de precisión no tienen en cuenta las relaciones debidas a estructuras espaciales en el cálculo de las ZM, ya que no han sido diseñados para detectar estas estructuras espaciales. Sin embargo existen formas de incorporar esta información espacial a los métodos multivariados a través del uso de variables sintéticas obtenidas a partir de la combinación de variables originales utilizando métodos geoestadístico clásicos. Dray *et al.* (2008), proponen un método de análisis que incorpora la información espacial previo al análisis multivariado. Este análisis espacial multivariado se basa en el índice de Moran para medir la dependencia o correlación espacial entre las observaciones, introduciendo una matriz de pesos espaciales estandarizada por fila (MULTISPATI) en el análisis. Este análisis, el cual es una generalización del análisis multivariado de correlación espacial de Wartenberg (1985), permite estudiar las relaciones entre las variables medidas (análisis multivariado) y su estructura espacial (autocorrelación). El objetivo de este trabajo es ilustrar la aplicación del

análisis de cluster utilizando el software MZA y el análisis MULTISPATI-PCA para la identificación de ZM de un lote en producción de granos.

Elementos del Trabajo y Metodología

Los datos recolectados en agricultura de precisión presentan la característica de estar georreferenciados, dado que se pretende identificar y explicar la variabilidad espacial de los rendimientos. Frecuentemente para poder explicar la variabilidad de la producción de un lote se miden varias variables para un mismo sitio en múltiples sitios del lote en producción. Esto genera bases de datos de gran volumen. En este sentido, las técnicas de análisis multivariado son implementadas con el objetivo de reducir la dimensión de la base de datos, revelar nuevas relaciones espaciales no evidentes e identificar patrones espaciales para la delimitación de ZM. En este trabajo se utilizan dos técnicas multivariadas para delimitar ZM, una basada en un análisis de cluster implementada en el software MZA y la otra basada en un análisis de componentes principales que incorpora la información de la estructura espacial *a priori* a través de una matriz de autocorrelación normalizada, esta técnica denominada MULTISPATI-PCA se encuentra implementada en el software libre R (www.r-project.org) como una función del paquete *ade4* (Chessel *et al.*, 2004). A partir del análisis de cluster se obtuvo una variable sintética conteniendo la clasificación de cada sitio a un determinado grupo o cluster, mientras que a partir del análisis MULTISPATI-PCA se obtuvieron como variables sintéticas las componentes principales espaciales (CPs). Posteriormente se realizó un análisis geoestadístico clásico ajustando semivariogramas sobre las variables sintéticas obtenidas de cada análisis multivariado. Con los valores predichos obtenidos de los semivariogramas, se construyeron mapas vía interpolación krigging (Schabenberger y Pierce, 2002), con el propósito de obtener una visualización gráfica de los resultados de las dos técnicas multivariadas aplicadas para la delimitación de ZM.

Estos procedimientos fueron implementados sobre un conjunto de datos provenientes de un lote en producción (73.21 ha) ubicado al sudeste bonaerense de la República Argentina, con información de 7576 sitios dentro del lote. Se compilaron valores georreferenciados de conductividad eléctrica aparente (CE_a) en dos profundidades 0-30 cm (CE_{a30}) y 0-90 cm (CE_{a90}) y elevación. Los valores de CE_a provienen de un sensor (Veris 3100, *Division of Geoprobe Systems*, Salina, KS) que utiliza el principio de la inducción electromagnética. El sensor Veris 3100 recorrió el lote en una serie de transectas paralelas espaciadas a intervalos de 15-20 m, debido a que una separación de más de 20 m genera errores de medición y la pérdida de información (Farahani *et al.*, 2007). El instrumento fue calibrado, según las instrucciones del fabricante, antes de la recolección de los datos. Los datos de CE_a fueron simultáneamente georreferenciados con un DGPS (Trimble R3, *Trimble Navigation Limited*, USA) con una exactitud de medición sub métrica y configurado para tomar la posición del satélite cada un segundo. Los datos de elevación (m) del terreno también se midieron con un DGPS y se procesaron para obtener una precisión vertical de aproximadamente 3-5 cm. Los datos fueron sometidos a procedimientos alternativos de depuración para disminuir el impacto de errores de las mediciones. En la instancia de pre-procesamiento de los datos, se usó como criterio de inclusión los datos que se encontraban entre la media ± 4 desvíos estándares. Se utilizó la transformación logaritmo natural sobre cada variable para linealizar las posibles relaciones no lineales que se puedan presentar dado que las técnicas multivariadas utilizadas se basan en correlaciones lineales. Finalmente, se realizó un análisis de co-inercia para obtener una medida de la magnitud de

la asociación entre los mapas generados por cada técnica multivariada, utilizando para ello el paquete ade4 del software R. Este análisis produce una medida global, a través del coeficiente RV, de la correlación entre las ordenaciones multivariadas. El coeficiente RV es alto cuando dos o más estructuras varían al mismo tiempo (alta correlación) y es bajo cuando su variación es independiente (Dray *et al.*, 2003). Para llevar a cabo el análisis de co-inercia, los mapas se convirtieron en matrices sitios×categoría, clasificando previamente los mapas en tres categorías. El consenso entre los mapas se estimó mediante el consenso entre estas matrices ($p < 0.005$).

Resultados

Las componente principales 1 y 2 (CPs1 y CPs2) resultantes del MULTISPATI-PCA explican un 99% de la variabilidad total de los datos. La variable con mayor importancia sobre la CPs1 para explicar la variabilidad espacial fue la elevación, ya que ésta presentó el mayor coeficiente (autovector) en relación al resto de las variables utilizadas en el análisis. Mientras que la CPs2 está representada principalmente por la CE_{a30}. Las variables CE_{a30}, CE_{a90} y elevación fueron introducidas en el software MZA para llevar a cabo el procedimiento de clasificación mediante el algoritmo fuzzy K-means y la distancia Euclídea. El exponente fuzzy se fijó en el valor convencional de 1.30 (Odeh *et al.*, 1992) y la clasificación se repitió variando la cantidad de ZM entre dos y seis ZM; luego, el software MZA utiliza dos índices para evaluar el número óptimo de ZM o cluster, el índice FPI y el índice NCE. El número óptimo de ZM es aquel que se obtiene cuando los índices de desempeño alcanzan su mínimo valor (Figura 1). Fridgen *et al.* (2004) indican que la concordancia de los dos índices es un indicador de la bondad de la clasificación. Dado que en nuestro ejemplo hay concordancia entre los índices, no serían necesarios llevar a cabo otro tipo análisis para verificar los resultados de la clasificación. En nuestro ejemplo ilustrativo, el número de ZM óptimo sugerido por el software MZA fue de tres.

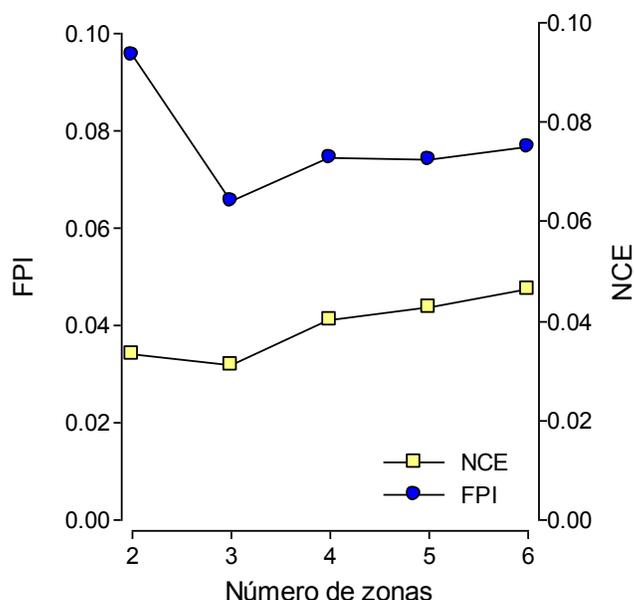


Figura 1. Índice de desempeño del grado de ambigüedad (Fuzziness Performance Index, FPI) y entropía de clasificación normalizada (Normalized Classification Entropy, NCE). La mejor clasificación se produce cuando el FPI y/o NCE obtienen los valores mínimos para el menor número de clases o zonas de manejo.

La Figura 2 muestra los mapas de las ZM potenciales del lote en estudio, realizados a partir de las variables sintéticas, para visualizar gráficamente los patrones de variabilidad espacial multivariados. Para el análisis de cluster se diferencian tres ZM como número óptimo ($k=3$), sin embargo se muestra también la clasificación obtenida habiendo definido dos ZM ($k=2$) (Fig. 2-d). Para obtener un valor de concordancia entre los patrones de variabilidad espacial obtenidos con MZA y MULTISPATI se realizó un análisis de co-inercia. El mapa obtenido a partir del análisis de cluster tuvo una mayor congruencia con el mapa proveniente la CPS2 del MULTISPATI, mientras que el coeficiente de correlación del análisis de cluster con la CPS1 fue inferior a pesar de que ésta explica una mayor proporción de la variabilidad total de los datos que la CPS2. Por otra parte, el consenso entre el análisis de cluster (cuando las ZM definidas fueron 2) y las dos primeras componentes espaciales, fue mayor.

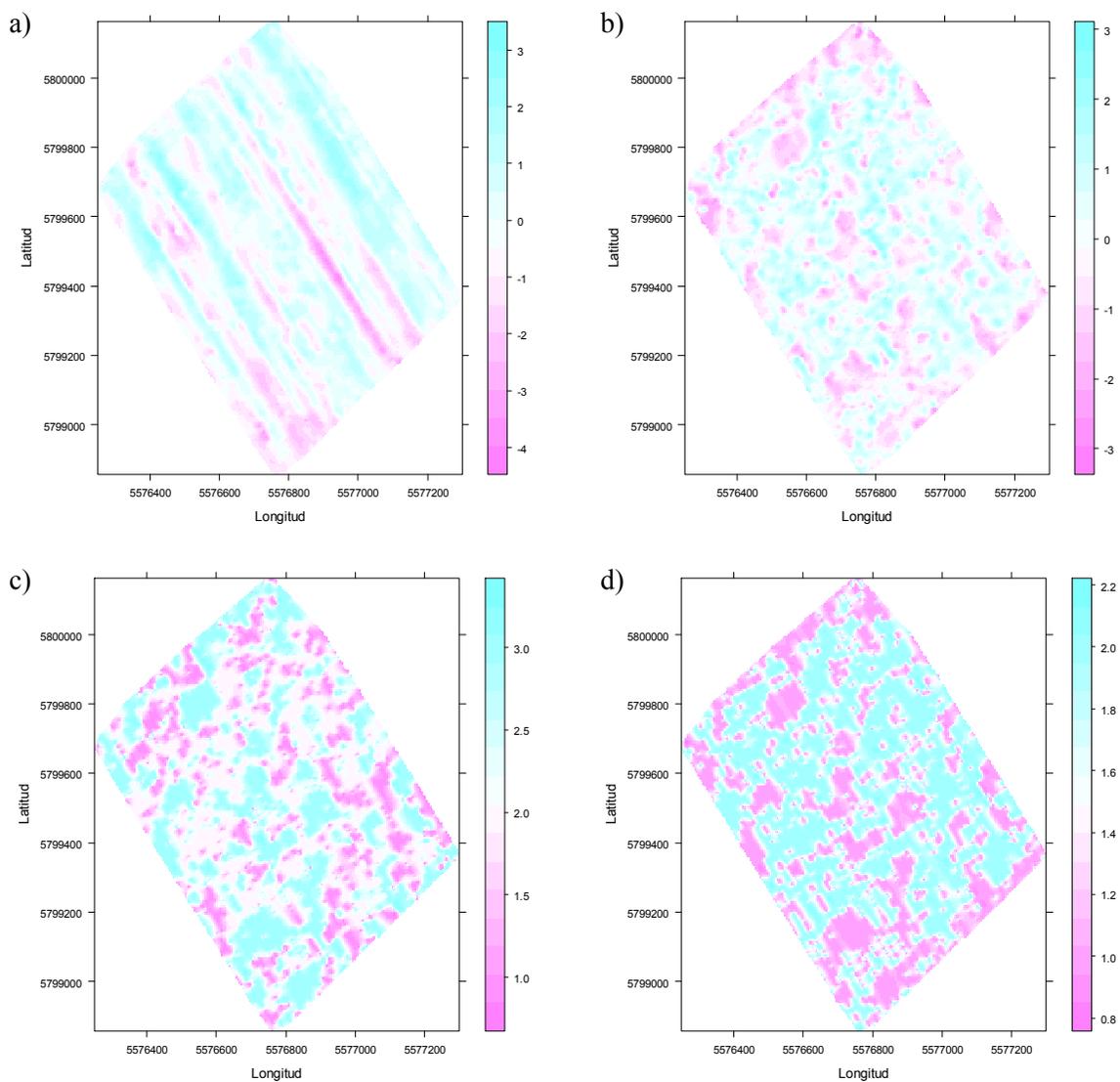


Figura 2. Mapas de variabilidad espacial obtenidos a partir de la interpolación kriging usando como variables la primera componente principal espacial (a), la segunda componente principal espacial (b), la clasificación obtenida del análisis de Cluster seleccionando tres zonas de manejo, $k=3$ (c) y seleccionando dos zonas de manejo, $k=2$ (d).

Discusión

Los resultados obtenidos a partir del análisis MULTISPATI estarían indicando que la elevación es una variable con importante inercia en la evaluación de la variabilidad espacial de los datos, sin embargo ésta significancia había sido débil cuando la información espacial no fue incorporada *a priori* en el análisis. De estos resultados, se desprende que incorporar la autocorrelación espacial *a priori* permite detectar relaciones subyacentes que no serían tenidas en cuenta para la clasificación de ZM si la estructura espacial fuera incorporada *a posteriori*. La identificación del número de ambientes o ZM y su tamaño está fuertemente relacionado a la variabilidad existente a nivel de lote (Saby *et al.*, 2009). El número de zonas está asociado al tipo de información utilizada y al método empleado para la clasificación. En la actualidad, existen disponibles una plétora de técnicas para la selección del número óptimo de ZM, provocando en el usuario, incertidumbre en cuanto a la eficiencia de cada una. Es por ello que consideramos necesario continuar investigando sobre el desempeño de los diferentes métodos para evaluar los resultados de la clasificación bajo distintos escenarios, esto implica la evaluación de diferentes variables en una escala espacial, es decir a través de varias campañas agrícolas, para evaluar también la estabilidad y continuidad espacial de las ZM en un lote en producción.

Agradecimientos

Los autores expresan sus agradecimientos a los productores, personal técnico y al Ing. Agr. Nahuel Peralta de la EEA INTA Balcarce, por su ayuda en la recopilación de los datos que se ofrecen en los campos donde se realizó el estudio. El presente trabajo es parte del trabajo de tesis de Mariano Córdoba para el cumplimiento de los requisitos del Doctorado en Ciencias Agropecuarias de la Universidad Nacional de Córdoba y del programa de becas de posgrado del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

Referencias

- [1] Bullock, D. S., Kitchen, N., Bullock D. G., 2007. Multidisciplinary Teams: A Necessity for Research in Precision Agriculture Systems. *Crop Sci.* 47, 1765–1769.
- [2] Chessel, D., Dufour, A.B., Thioulouse, J., 2004. The ade4 package I: one-table methods. *R news* 4 (1), 5–10.
- [3] Corwin, D.L., Lesch, S.M., Shouse, P.J., Soppe, R., Ayars, J.E., 2003. Identifying soil properties that influence cotton yield using soil sampling directed by apparent soil electrical conductivity. *Agron. J.* 95, 352–364.
- [4] Corwin, D.L., Lesch, S.M., 2010. Delineating Site-Specific Management Units with Proximal Sensors. In: *Geostatistical Applications for Precision Agriculture*. Oliver M.O. (ed), Springer, Netherlands, pp.139–165.
- [5] Dray, S., Chessel, D., Thioulouse, J., 2003. Procrustean co-inertia analysis for the linking of multivariate datasets. *Ecoscience* 10, 110–119.
- [6] Dray, S., Said, S., Debias, F., 2008. Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. *J. Veg. Sci.* 19, 45–56.
- [7] Farahani, H.J., Flynn, R.L., 2007. Map Quality & Zone Delineation as affected by Width of Parallel Swaths of Mobile Agricultural Sensors. *Precision Agriculture* 96, 151–159.
- [8] Fraisse, C., Sudduth, K., Kitchen, N., 2001. Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. *Transactions of the ASAE.* 44, 155–166.
- [9] Fridgen, J.J., Kitchen, N.R., Sudduth, K.A., Drummond, S.T., Wiebold, W.J., Fraisse, C.W., 2004. Management Zone Analyst (MZA): Software for Subfield Management Zone Delineation. *Agron. J.* 96,100–108.
- [10] Jaynes, D.B., Colvin, T.S., Kaspar, T.C., 2005. Identifying potential soybean management zones from multi-year yield data. *Comp. Electron. Agric.* 46,309-327
- [11] Johnson, C.K., Mortensen, D.A., Wienhold, B.J., Shanahan, J.R., Doran, J.W., 2003. Site-specific management zones based on soil electrical conductivity in a semiarid cropping system. *Agron. J.* 95, 303–315.

- [12] Kitchen, N.R., Sudduth, K.A., Myers, D.B., Drummond, S.T., Hong S.Y., 2005. Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. *Comp. Electron. Agric.* 46, 285–308.
- [13] Kühn, J., Brenning, A., Wehrhan, M., Koszinski, S., Sommer, M., 2009. Interpretation of electrical conductivity patterns by soil properties and geological maps for precision agriculture. *Precision Agriculture*. 10,490–507.
- [14] Moral, F.J., Terrón, J.M., 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, 335–343.
- [15] Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1992. Soil pattern recognition with fuzzy c-means: application to classification and soil-landform interrelationship. *Soil Sci. Soc. Am. J.* 56, 505–516.
- [16] Okabe, A., Boots, B., Sugihara, K., Chiu, S., 2000. *Spatial tessellations: concepts and applications of Voronoi diagrams*. John Wiley and Sons Ltd., England.
- [17] Ortega, R. A., Santibañez, O. A., 2007. Determination of management zones in corn (*Zea mays* L.) based on soil fertility. *Comp. Electron. Agric.* 58, 49–59.
- [18] Schabenberger, O., Pierce F.J., 2002. *Contemporary statistical models for the plant and soil sciences*. Taylor and Francis. CRC Press, Boca Raton, Florida.
- [19] Saby, N.P.A., Thioulouse, J., Jolivet, C.C., Ratié, C., Boulonne, L., Bispo, A., Arrouays, D., 2009. Multivariate analysis of the spatial patterns of 8 trace elements using the French soil monitoring network data. *Science of the Total Environment*, 407, 5644–5652.
- [20] Wartenberg, D., 1985. Multivariate spatial correlation: a method for exploratory geographical analysis. *Geogr. Anal.* 17, 263–283.

Datos de Contacto:

Mariano Augusto Córdoba. Cátedra de Estadística y Biometría. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba. Av. Valparaíso s/n. Ciudad Universitaria. CC 509. CP 5000. Córdoba. marianoacba@agro.unc.edu.ar.