

Geoestadística multivariada en agricultura de precisión

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

Nuevas tecnologías de agricultura de precisión permiten recolectar información de múltiples variables en gran cantidad de sitios georreferenciados dentro de lotes en producción. El estudio de la variabilidad espacial de las propiedades del suelo y su relación con la variabilidad del rendimiento de los cultivos, es un paso crucial para la agricultura de precisión. La variabilidad espacial de cada variable estudiada es visualizada a través de mapas obtenidos por interpolación geoestadística para cada variable. En este trabajo se ilustra la elaboración de mapas de variabilidad espacial multivariada, es decir teniendo en cuenta las múltiples variables registradas sobre cada sitio simultáneamente. Estos se construyen a partir de variables sintéticas resultantes de la combinación de las variables originales. Las variables sintéticas maximizan la representación de la variabilidad total de los datos, tanto de suelo como de rendimiento. Se evaluó el consenso entre mapas de variabilidad espacial multivariados basados en los algoritmos del análisis de componentes principales, correlaciones canónicas, árboles de regresión-clasificación y análisis de cluster, como herramientas para la obtención de variables sintéticas. Los mapas de variabilidad, basados en variables producidas por algoritmos multivariados, resultaron potentes respecto a su capacidad para identificar ambientes homogéneos dentro de los lotes.

Abstract

New precision agriculture technologies allow collecting information from several variables at many georeferenced sites within crop fields. The study about spatial variability of soil properties and their relationship with crop yield variability is a crucial step in precision agriculture. The spatial variability of each variable is usually viewed through maps generated by geostatistical interpolation. This paper illustrates the mapping of multivariate spatial variability, i.e. taking into account the multiple variables recorded at each site simultaneously. Such maps are built from synthetic variables resulting from the combination of the original ones. These synthetic variables maximize the representation of the total variability of soil and yield data. Principal component, canonical correlations cluster and classification-regression trees based-maps were compared. Multivariate maps were powerful in their ability to identify homogeneous environments within field.

Palabras Clave: Mapas de variabilidad, manejo sitio-específico, variables sintéticas.

Introducción

Desde los comienzos de la agricultura, la heterogeneidad de los rendimientos en un lote hace pensar que el manejo del cultivo en un área determinada del lote debiera ser realizada en función de sus características particulares (Bullock *et al.*, 2007). Sin embargo, aún son escasos los estudios de variabilidad a escala fina (intralote) tanto de las propiedades edáficas y topográficas como de la productividad, por ello la mayoría de los lotes reciben un manejo uniforme. Las nuevas tecnologías en maquinarias agrícolas asociadas a la agricultura de precisión (AP) proporcionan la oportunidad de medir con mayor precisión la variabilidad en el rendimiento y en las variables relacionadas (Roel y Plant, 2004). La medición georreferenciada de éstas se ha convertido en una herramienta tecnológica invaluable ya que permite identificar las propiedades físico-químicas del suelo y determinar la variación espacial de las mismas. El estudio de la variabilidad espacial de las propiedades del suelo y su relación con la distribución espacial del rendimiento de los cultivos dentro del lote, permite la delimitación de subregiones que expresan una

combinación relativamente homogénea de factores de rendimiento, tales como la textura, la topografía y el nivel de nutrientes. Tradicionalmente, las covariaciones o correlaciones espaciales de las propiedades del sitio y el rendimiento de los cultivos han sido abordadas a través de modelos geoestadísticos clásicos. La geoestadística utiliza variables medidas en un lote con referencia espacial de cada punto o sitio de muestreo que se supone pertenece a un campo aleatorio continuo. Se dice que existe autocorrelación espacial positiva si observaciones de dos sitios son más similares entre sí en la medida en que éstos se encuentren más cercanos en el espacio (Cressie, 1993). Si bien, con la actual AP, en cada sitio se pueden evaluar múltiples variables, el abordaje clásico del análisis de autocorrelación espacial se realiza variable a variable, es decir de manera univariada (Schabenberger y Pierce, 2002). Los mapas obtenidos por interpolación geoestadística univariada, han sido la principal herramienta utilizada para visualizar la variabilidad espacial de distintas variables relacionadas a los patrones espaciales de productividad de los cultivos. Dado que en cada sitio del lote se recolecta un conjunto de variables y que el rendimiento depende de todos ellos y sus interacciones o correlaciones, es que proponemos abordar el estudio de variabilidad espacial dentro del lote desde un enfoque multivariado. Para ello se utilizan mapas derivados de variables sintéticas que resumen la variabilidad. Identificando patrones de variación espacial conjunta, se pretende identificar sitios o ambientes dentro del lote donde los insumos agrícolas podrían aplicarse de manera distinta. Este manejo sitio-específico mejora la eficiencia en el uso de los recursos, contribuyendo de manera directa a la protección del medio ambiente realizando un uso adecuado de agroquímico y potenciando el rendimiento de los cultivos de manera sustentable en el tiempo (Dinnes *et al.*, 2002). Además, la zonificación permite identificar áreas homogéneas que pueden ser útiles en el diseño de muestreos de características que se ven afectados por la heterogeneidad espacial del lote. Las aproximaciones metodológicas basadas en el análisis multivariado si bien han sido ampliamente usadas en ensayos agrícolas, sólo recientemente se están usando para el tratamiento de datos georreferenciados y muchas de ellas, no han sido desarrolladas para manejar datos espaciales a una escala fina. Un tipo de algoritmo computacional difundido específicamente para la delimitación de ambientes, es el análisis de cluster o conglomerado que consiste en la identificación de grupos con la particularidad de que las observaciones que conforman cada grupo son más similares entre sí. En el análisis de cluster los grupos (clases) 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). Roel y Plant (2004) utilizan el análisis de cluster para discernir los patrones espacio-temporales en el rendimiento de grano y a continuación otro tipo de algoritmo particionante como los árboles de clasificación y regresión (CART) para determinar los factores subyacentes en la distribución de los clusters. El algoritmo CART es un método estadístico no paramétrico que particiona de forma recursiva el espacio multidimensional definido por las variables explicativas en subconjuntos que son lo más homogéneas posible en términos de los valores de la variable de respuesta (Breiman *et al.*, 1984). A diferencia de los modelos paramétricos, que van a descubrir una única estructura dominante en los datos, este método está diseñado para trabajar con datos en los que podría haber múltiples estructuras. Esta característica lo hace sumamente flexible en AP donde muchas veces dentro de un mismo lote pueden existir una serie de variables explicando la variabilidad de rendimiento a diferentes niveles y zonas (Roel y Plant, 2004). Por otro lado existe la alternativa de usar

análisis de componentes principales (ACP), este se ha utilizado en la AP para analizar los efectos de las propiedades del suelo y los atributos del terreno sobre el rendimiento (Kaspar *et al.*, 2004; Moral *et al.*, 2010), pero raramente combinando datos de suelo y rendimiento con propósitos de manejo. El ACP es una técnica multivariada que permite identificar las variables que explican la mayor parte de la variabilidad total contenida en los datos originales, explorar sus correlaciones y reducir las dimensiones (ejes) de análisis de la información creando nuevas variables no correlacionadas (variables sintéticas). Finalmente otra técnica basada en correlaciones entre variables es el análisis de correlaciones canónicas (ACC). El ACC permite cuantificar la magnitud de la relación lineal entre dos grupos de variables métricas, unas consideradas como variables dependientes y otras como independientes. En la AP el ACC se ha utilizado en el manejo sitio-específico de nemátodos de suelo (Ortiz *et al.*, 2007), el mapeo del drenaje de suelo a nivel de lote (Liu *et al.*, 2008) y para estudiar la asociación entre las características de campo y rendimiento del cultivo de soja (Martin *et al.*, 2005). No obstante, no ha sido utilizado para construir mapas de variación conjunta. Si bien existen antecedentes en la implementación de estas técnicas multivariadas en problemas de AP, son raras las evaluaciones de los resultados obtenidos vía la aplicación simultánea de estos algoritmos sobre un mismo caso de estudio (lote). El objetivo de nuestro estudio es evaluar la congruencia de mapas obtenidos aplicando diferentes algoritmos multivariados para analizar la variación conjunta de datos de suelo y de rendimiento con el propósito de identificar áreas homogéneas en sentido estadístico multivariado o estadístico a escala de lote.

Elementos del Trabajo y Metodología

Las técnicas geoestadísticas clásicas, comúnmente usadas en AP para describir variación espacial, se aplican en sentido univariado resultando difícil la interpretación de la variabilidad conjunta causadas por las relaciones entre variabilidad del rendimiento y de propiedades de suelo. Para comprender la relación entre estas variables el análisis de correlación y de regresión lineal múltiple han sido implementados (Kravchenko y Bullock, 2000). Los resultados de estos análisis generalmente no son satisfactorios, aún cuando se consideran términos polinómicos e interacciones en el modelo (Kitchen *et al.*, 1999). Las técnicas multivariadas, representan un nuevo enfoque para hacer frente a la necesidad de analizar múltiples correlaciones (Kaspar *et al.*, 2004). La hipótesis que sustenta nuestra propuesta es que tales técnicas facilitan la interpretación de las complejas relaciones, reduce la dimensión de la base de datos, permite detectar estructuras en los datos y revelar nuevas relaciones espaciales no evidentes cuando las variables se analizan individualmente. Los métodos utilizados en este trabajo son todas técnicas multivariadas. Estas fueron implementadas sobre un conjunto de datos provenientes de un lote en producción 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}), elevación, rendimiento de soja cosecha 2006 (RtoSj) y rendimiento de trigo cosecha 2006 (RtoTg) en cada sitio. 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. Las observaciones fueron georreferenciadas con un GPS (trimble, geoexplorer 2005) cada 4 m de distancia dentro de transectas separadas aproximadamente 20 m. La latitud, longitud y elevación (m) del terreno se midieron con un DGPS (Trimble R3, Trimble Navigation Limited, USA). 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 $\pm 4DE$. Para cuantificar el rendimiento en grano de los cultivos se utilizaron monitores de rendimiento acoplados a los equipos de cosecha. Debido a la mayor resolución espacial de la variable rendimiento respecto a las variables de suelo, se llevó a cabo una interpolación del rendimiento mediante el método del vecino natural ponderado que estima el valor promedio de rendimiento ponderado por el área que conforman los puntos adyacentes, dicha área o vecindario se obtiene por la diferencia de los conjuntos de polígonos de Thiessen (Okabe *et al.*, 2000). Posterior a la interpolación se realizó la extracción del valor de rendimiento para cada punto georreferenciado de las propiedades de suelo, quedando de esta manera todas las variables con la misma resolución espacial. Se utilizó la transformación logaritmo natural sobre cada variable para linealizar las posibles relaciones no lineales que se puedan presentar dado que la mayoría de las técnicas multivariadas utilizadas se basan en correlaciones lineales. Esta transformación permite además aproximarlas a una distribución simétrica.

Los análisis multivariados realizados sobre estos datos incluyeron: 1) ACP sobre una matriz $n \times p$ donde n es la cantidad de sitios de muestreo en el lote y p representa las variables que se incluyeron en el análisis reteniéndose para ambos lotes las dos primeras componentes principales (CP1, CP2), 2) ACC usando variables de suelo y de rendimiento para conformar los grupos, obteniéndose como variables sintéticas las variables canónicas de suelo y de rendimiento correspondientes a la primera relación canónica significativa, 3) algoritmo CART para obtener valores umbral de las variables de suelo que predicen rendimientos para el cultivo de soja y trigo. Los árboles de clasificación (CT) son usados para predecir variables categorizadas para lo cual el algoritmo de partición se aplicó sobre tres categorías de rendimiento previamente definidas según los percentiles 33 y 66 (rendimiento bajo, rendimiento medio y rendimiento alto). El árboles de regresión (RT) que a diferencia del CT permiten predecir variables respuestas continuas se aplicó usando la variable rendimiento como continua. Las variables sintéticas obtenidas tanto en el CT y RT corresponden a la clasificación de los sitios según los umbrales de las variables regresoras. 4) También se realizó un análisis de cluster de los sitios de muestreo utilizando el algoritmo de agrupamiento no jerárquico k-means, reteniéndose como variable sintética la variable categorizada que indica el grupo al que fue clasificado cada sitio. Se definieron para la clasificación tres grupos *a priori*.

Posteriormente, se realizó un análisis geoestadístico clásico ajustando semivariogramas sobre las variables sintéticas obtenidas de los análisis multivariados. Con los valores predichos obtenidos de los semivariogramas, se construyeron mapas vía interpolación kriging universal recomendado en casos donde la variable no cumpla el supuesto de estacionariedad y se caracterice por exhibir una tendencia (Schabenberger y Pierce, 2002). Para comprobar el supuesto de estacionariedad se ajustó un modelo de regresión lineal múltiple en los que se utilizó como variables independientes las coordenadas planas (longitud y latitud) y su interacción.

Finalmente, para obtener una medida de la magnitud de la asociación entre todos los pares de mapas obtenidos de los distintos análisis multivariados, se estimaron los coeficientes de consenso vía análisis de co-inercia (Dray *et al.*, 2003). Este análisis produce una medida global (RV) de la correlación entre las ordenaciones multivariadas. El coeficiente RV es alto, cuando dos o más estructuras varían al mismo tiempo (también cuando varía inversamente), y es bajo cuando su variación es independiente. Para llevar a cabo el análisis

de co-inercia, los mapas se convirtieron en matrices sitios×categoría, clasificando previamente los mapas que fueron construidos con variables continuas en tres categorías. El consenso entre los mapas se estimó mediante el consenso entre estas matrices ($p < 0.005$). Para la obtención de las variables sintéticas mediante algoritmos multivariados se utilizó el software InfoStat (Di Rienzo *et al.*, 2011) mientras que para el análisis geoestadístico y de co-inercia se utilizó el mismo software pero utilizando una interfase que lo conecta con el software R (www.r-project.org). Las librerías utilizadas para estos análisis fueron gstat (Pebesma, 2004) y ade4 (Dray *et al.*, 2007) respectivamente.

Resultados

En las Figura 1 y 2 se representan gráficamente los resultados producidos por los algoritmos multivariados. El biplot del ACP sugiere, a nivel de la primera componente principal (CP1), alta correlación negativa entre la variable CE_{a30} respecto al rendimiento; valores menores de CE_{a30} implican mayores rendimientos, tanto en soja (RtoSj) como trigo (RtoTg) (Figura 1a). La CE_a es una de las variables de suelo comúnmente utilizada en AP, su medición georreferenciada caracteriza la variabilidad espacial incluso de otras propiedades edáficas (Corwin *et al.*, 2005) que determinan la productividad de un cultivo, tales como humedad del suelo, CIC, MO, textura y contenido de sales. Dichas características hacen que la CE_a sea considerada un buen predictor del rendimiento en grano en distintos sitios del lote de producción (Johnson *et al.*, 2003). La elevación es otra propiedad topográfica comúnmente utilizada para evaluar la variabilidad de los rendimientos (Kaspar *et al.*, 2004). Esta afecta directamente el crecimiento y desarrollo de los cultivos por la acumulación de agua en diferentes partes del terreno e indirectamente por la erosión y deposición del suelo (Pennock y Jong, 1987).

La primera variable canónica del ACC es definida principalmente para el grupo de variables de suelo por CE_{a30} y para el de rendimiento por RtoSj. En el gráfico de dispersión se observa una relación lineal negativa entre ambos grupos de variables (Figura 1b).

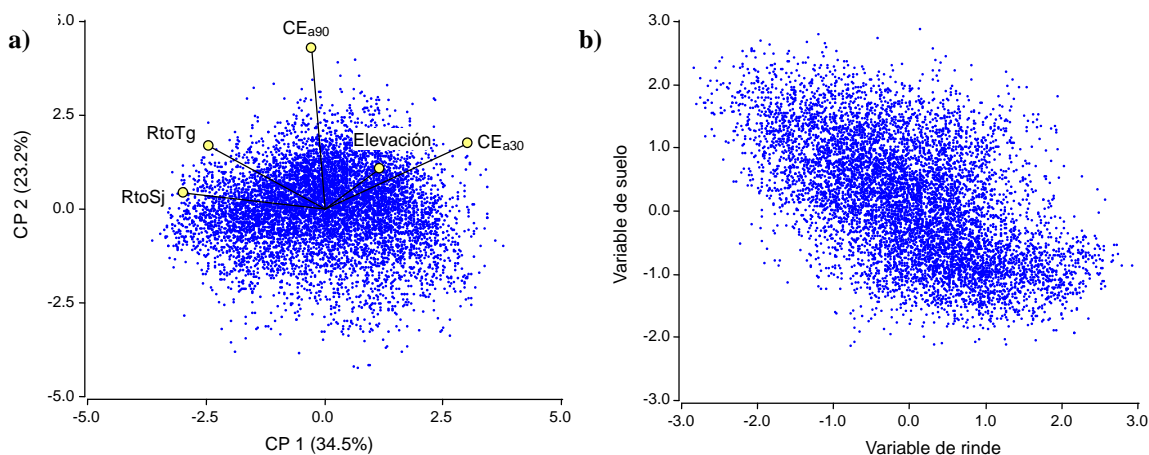


Figura 1. a) Biplot de la primera y segunda componente principal del ACP: variabilidad de 7.576 sitios del lote, considerando la conductividad eléctrica aparente a 30 cm (CE_{a30}) y 90 cm (CE_{a90}), elevación, rendimiento de soja (RtoSj) y rendimiento de trigo (RtoTg). b) Diagrama de dispersión de las primeras variables canónicas de los datos de rendimiento y suelo.

El algoritmo CART para la variable RtoSj identificó a la CE_{a30} como la primer variable que permitió generar dos grupos de puntos (nodos), valores umbrales de CE_{a30} menores a 24.53

mS m^{-1} predicen altos rendimientos promedio. Los umbrales estimados fueron similares con ambos árboles (CT y RT). A su vez dentro del grupo de observaciones que tuvieron un menor rendimiento promedio, valores de CE_{a90} menores a 28.78 mS m^{-1} predijeron valores de rendimiento promedio bajos en ambos árboles (Figura 2).

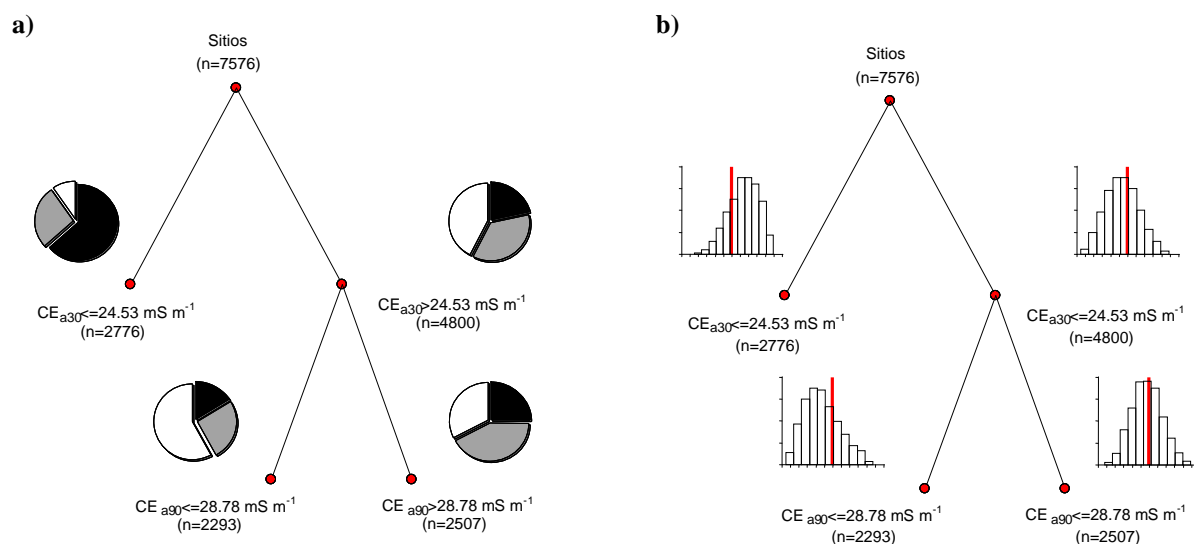


Figura 2. a) Árbol de clasificación particionando 7576 sitios del lote de acuerdo a la relación entre las categorías de rendimiento de los sitios y las variables del suelo representada por la conductividad eléctrica aparente a los 30 cm (CE_{a30}) y 90 cm (CE_{a90}). Se muestran los valores de umbral y la distribución de la categoría de rendimiento en cada nodo (negro: rendimiento alto, gris: rendimiento medio, blanco: rendimiento bajo). b) Árbol de regresión particionando los mismos sitios de acuerdo a la relación entre el rendimiento del sitio (como variable continua) y las variables del suelo representadas también por la CE_a .

A partir de los resultados de cada técnica obtenidas por distintos algoritmos multivariados, se confeccionaron mapas para las principales variables sintéticas. Obteniéndose 6 mapas (Figuras 3 y 4): 1) “CP1” (resultante del ACP), 2) “CanSu” (para el grupo de variables de suelo del ACC), 3) “CanRi” (para el grupo de variables de rendimiento del ACC), 4) “RTCT-Sj” (RT y CT del cultivo de soja); como los umbrales fueron similares se realizó un único mapa a partir de los resultados obtenidos del RT, 5) “RTCT-Tg” (RT y CT del cultivo de trigo); al igual que para los resultados de rendimiento de soja los umbrales fueron similares entre RT y CT, por ello que se realizó un único mapa con RT y 6) “k-means”. Los coeficientes de correlación obtenidos a partir del análisis de co-inercia entre todos los pares de mapas se presentan en la Tabla 1.

Tabla 1. Coeficientes de correlación entre mapas elaborados a partir de variables sintéticas.

	CP1	CanSu	CanRi	RTCT-Sj	RTCT-Tg	k-means
CP1	1					
CanSu	0.348	1				
CanRi	0.423	0.151	1			
RTCT-Sj	0.261	0.406	0.152	1		
RTCT-Tg	0.274	0.445	0.150	0.615	1	
k-means	0.506	0.152	0.121	0.180	0.152	1

Los coeficientes de la Tabla 1, muestran que la mayor congruencia se obtuvo entre los mapas RTCT-Sj y RTCT-Tg ($RV=0.615$). El mapa de las variables sintéticas resultantes del ACP mostró un alto grado de consenso con k-means ($RV=0.506$), CanRi ($RV=0.423$) y CanSu ($RV=0.348$). Esto último evidencia la naturaleza multivariada de las variables sintéticas obtenidas con el ACP (CP1), la cual se correlaciona con ambos rendimientos (negativamente) y con CE_{a30} (positivamente). Los mapas obtenidos con el algoritmo CART para los rendimientos trigo y soja se correlacionan con CanSu ($RV=0.445$ y 0.406 respectivamente). La ventaja de utilizar métodos como CART es que estos no se ven afectados por la existencia de relaciones no lineales o relaciones que no se hayan podido linealizar con la transformación realizada en el pre-procesamiento de los datos. Además, este algoritmo también representa un enfoque simplificado para identificar los umbrales de los parámetros relacionados con el rendimiento alcanzable.

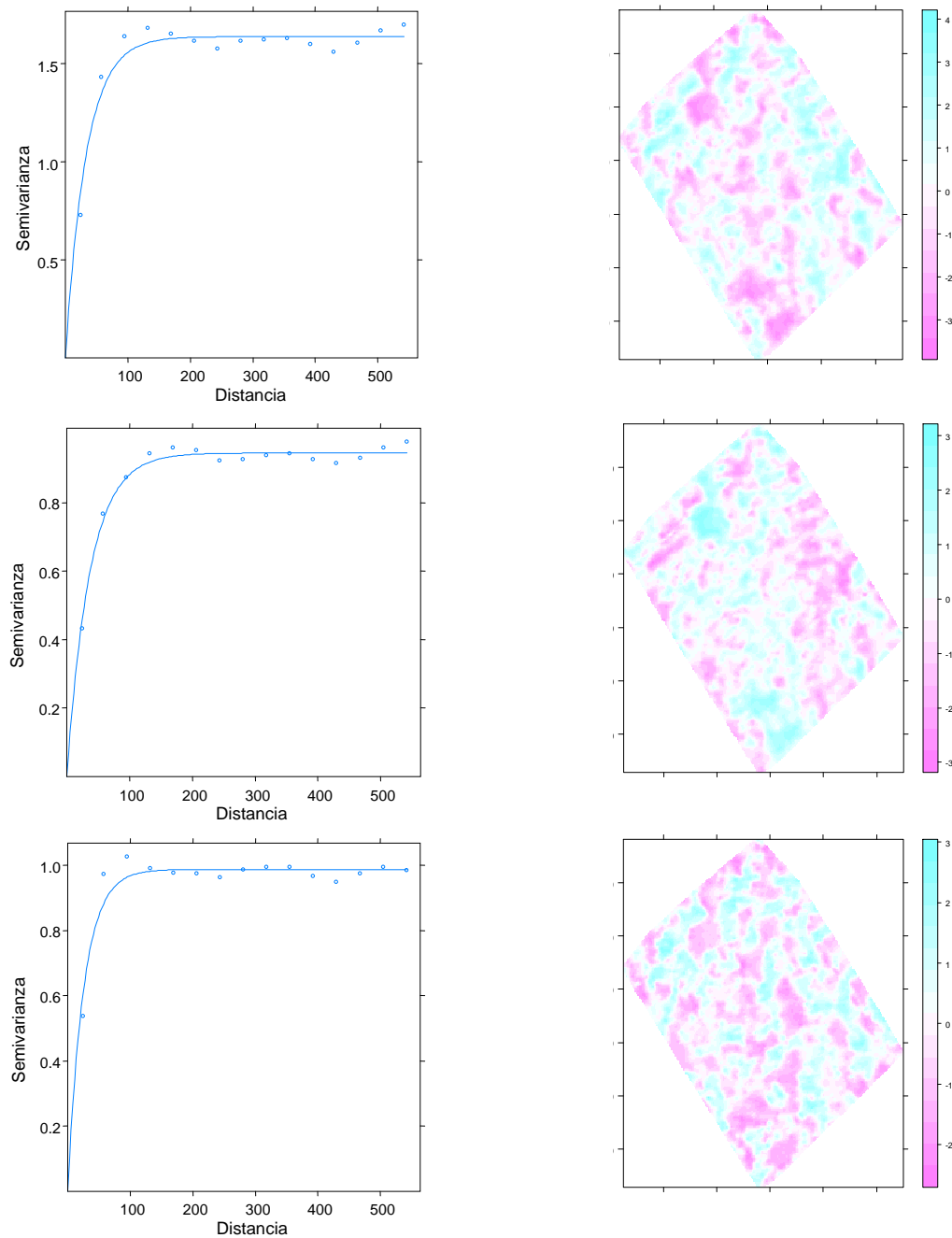


Figura 3. Semivariogramas empíricos y experimentales ajustados (izquierda) y mapas de variabilidad espacial multivariada intralote obtenidos a partir de variables sintéticas (derecha). Arriba: CP1 del ACP de los datos de rendimiento y el suelo, medio: primer variable canónica de los datos de rendimiento (CanRi); abajo: primer variable canónica de los datos de suelo (CanSu).

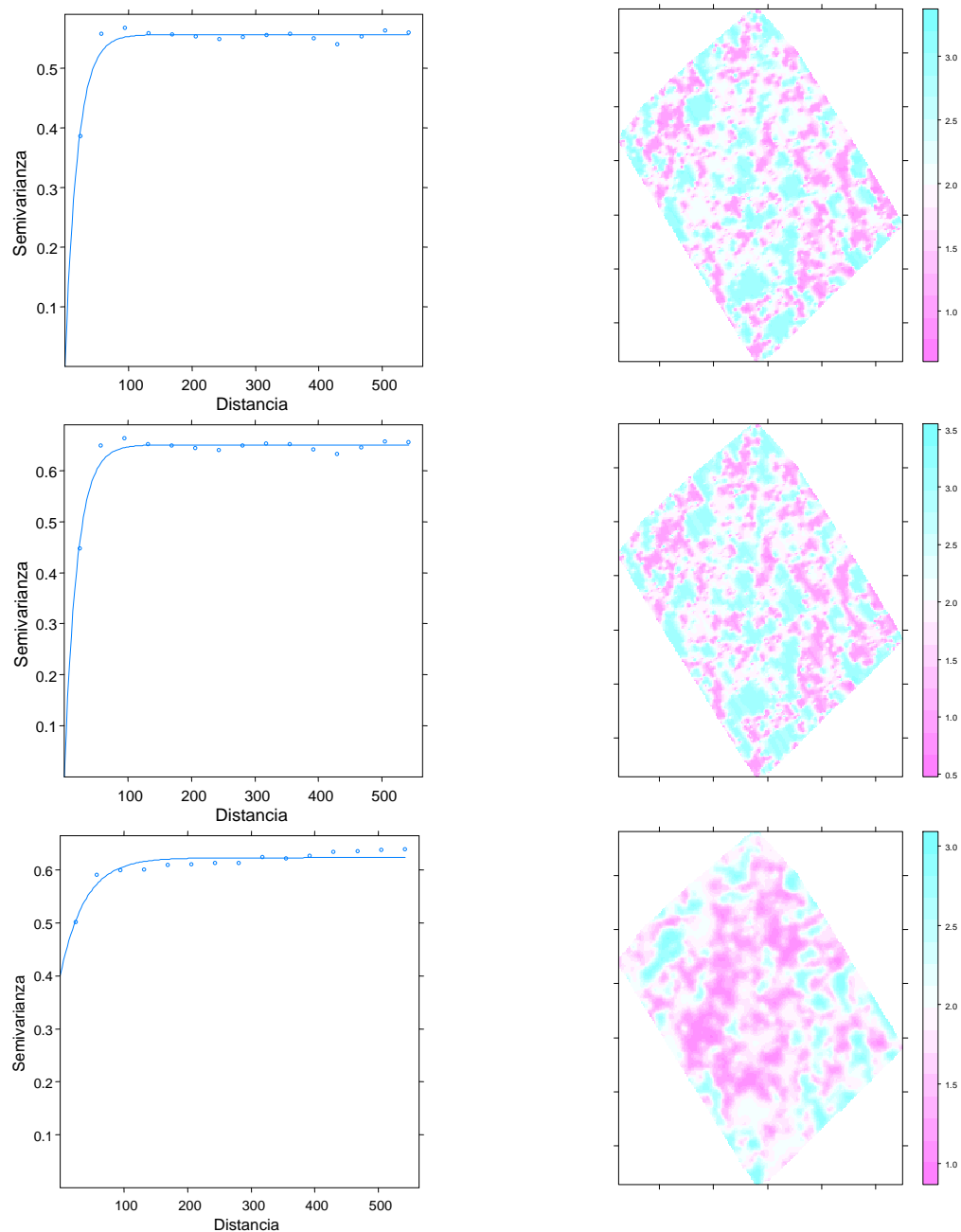


Figura 4. Semivariogramas empíricos y experimentales ajustados (izquierda) y mapas de variabilidad espacial multivariada intralote obtenidos a partir de variables sintéticas (derecha). Arriba: rendimiento de soja predichos a partir de variables del suelo por RT (RTCT-Sj), medio: rendimientos de trigo predichos a partir de variables del suelo por RT (RTCT-Tg); inferior: análisis de cluster (k-means).

Discusión

La identificación de zonas homogéneas en sentido estadístico multivariado a través de mapas confeccionados a partir de variables sintéticas obtenidas desde técnicas multivariadas ayuda a manejar simultáneamente datos de suelo con datos de rendimientos obtenidos bajo diferentes condiciones ambientales y de manejo. Además el uso de algoritmos multivariados permite identificar las variables que explican la variabilidad espacial global. Estas herramientas permiten la visualización de resultados multivariados

ayudando a comprender las relaciones complejas entre las variables y su impacto en la identificación de zonas homogéneas dentro del lote. La identificación de estas zonas es importante para realizar un muestreo zonificado que permite reducir la cantidad de muestras sin perder la información de la variabilidad subyacente. Partiendo del supuesto de que la agrupación de sitios en zonas homogéneas en un sentido estadístico multivariado reducirá la variabilidad dentro de la zona, todos los análisis que se proponen brindan una oportunidad para la identificación de zonas de manejo intralote y, potencialmente, la aplicación de manejo sitio específico que permitirá maximizar la producción de cultivos a través de todo el lote.

El número de zonas de manejo depende de la medida de sensibilidad deseada y el nivel de variabilidad del lote (Li Yan *et al.*, 2007). Para este estudio, la separación en tres zonas demostró ser un buen compromiso entre la sensibilidad y los patrones de variabilidad visualmente perceptible del suelo y rendimiento. El análisis de cluster no jerárquico (k-means) obtuvo los coeficientes de co-inercia más bajos respecto a la congruencia con los otros mapas excepto con CP1. Sin embargo, este análisis es frecuentemente utilizado en AP debido a que es una técnica simple de aplicar y se encuentra disponible en la mayoría de los programas, convirtiéndola en una de las más frecuentes para la identificación de zonas o sitios homogéneos desde una perspectiva multivariada.

La estabilidad temporal de las zonas homogéneas definidas está asociada a las características particulares de cada lote productivo, con resultados muy dispares, siendo relativamente estables en algunos casos (Jaynes *et al.*, 2005) y en otros muy variables, asociado principalmente al tipo de información utilizada para su delimitación (Taylor *et al.*, 2003) como así también a la interacción de los factores edáficos, climáticos y antrópicos. En este trabajo, la variación espacial de propiedades de suelo y rendimientos de una sola campaña agrícola se utilizaron para delimitar zonas de manejo, la variabilidad temporal o la estabilidad de estas variables no se ha tenido en cuenta. Además, la continuidad espacial de las zonas de manejo no sólo está asociada con las variaciones espacio-temporales de las variables utilizadas, sino que también se asocia con su escala espacial (Li Yan *et al.*, 2007). Por lo tanto, será importante utilizar las diferentes variables registradas en varias campañas u años agrícolas para definir zonas de manejo de acuerdo a la variabilidad espacial y estabilidad temporal de los atributos del suelo y rendimiento de los cultivos, utilizando datos con diferentes escalas espaciales, para evaluar la continuidad espacial de las zonas de manejo.

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] Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and regression trees. Chapman and Hall, New York.
- [2] Bullock, D. S., Kitchen, N., Bullock D. G., 2007. Multidisciplinary Teams: A Necessity for Research in Precision Agriculture Systems. Crop Sci. 47, 1765–1769.
- [3] Corwin, D.L., Lesch S.M., 2005. Characterizing soil spatial variability with apparent soil electrical conductivity. Part I: Survey protocols. Comp. Electron. Agric. 46, 103–133.
- [4] Cressie, N.A.C., 1993. Statistics for Spatial Data. Revised Ed. John Wiley and Sons, New York.
- [5] Dinnes, D.L., Karlen, D.L., Jaynes, D.B., Kaspar, T.C., Hatfield, J.L., Colvin, T.S., Cambardella, C.A.,

2002. Nitrogen management strategies to reduce nitrate leaching in tile-drained Midwestern soils. *Agron. J.* 94, 153–171.
- [6] Di Rienzo, J.A., Casanoves, F., Balzarini, M.G., Gonzalez, L., Tablada, M., Robledo, C.W., InfoStat versión 2011. Grupo. InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. URL <http://www.infostat.com.ar>.
- [7] Dray, S., Dufour, A.B., Chessel, D., 2007. The ade4 package-II: Two-table and K-table methods. *R News.* 7, 47–52.
- [8] Dray, S., Chessel, D., Thioulouse, J., 2003. Procrustean co-inertia analysis for the linking of multivariate datasets. *Ecoscience* 10, 110–119.
- [9] 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.
- [10] 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.
- [11] Kaspar, T.C., Pulido, D.J., Fenton, T.E., Colvin, T.S., Karlen, D.L., Jaynes, D.B., Meek, D.W., 2004. Relationship of corn and soybean yield to soil and terrain properties. *Agron. J.* 96: 700–709.
- [12] Kitchen, N. R., Sudduth, K. A., Drummond, S. T., 1999. Electrical conductivity as a crop productivity measure for claypan soils. *Journal of Production Agriculture.* 12, 607–617.
- [13] Kravchenko, A.N., Bullock, D.G., 2000. Correlation of corn and soybean grain yield with topography and soil properties. *Agron. J.* 92, 75–83.
- [14] Li, Y., Shi, Z., Li, F., Li, H.Y., 2007. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. *Comp. Electron. Agric.* 56, 174–186.
- [15] Liu, J., Pattey, E., Nolin, M.C., Miller, J.R., Ka, O., 2008. Mapping within-field soil drainage using remote sensing, DEM and apparent soil electrical conductivity. *Geoderma* 143, 261–272.
- [16] Martin, N.F., Bollero, G.A., Bullock, D.G., 2005. Associations between field characteristics and soybean plant performance using canonical correlation analysis. *Plant Soil* 273, 39–55.
- [17] 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.
- [18] 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.
- [19] Ortiz, B.V., Sullivan, D.G., Perry, C., Vellidis, G., Seymour, L., Rucker, K., 2007. Delineation of management zones for site specific management of parasitic nematodes using geostatistical analysis of measured field characteristics. In: Stafford, J., and Werner, A. (Eds.), *Proceedings of the sixth European conference on precision agriculture (6ECPA)*. Skiathos, Greece. 3-6 Jun. 2007.
- [20] Okabe, A., Boots, B., Sugihara, K., Chiu, S., 2000. *Spatial tessellations: concepts and applications of Voronoi diagrams*. John Wiley and Sons Ltd., England.
- [21] Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers and Geosciences*, 30, 683–691.
- [22] Pennock, D. J., De Jong, E., 1987. The influence of slope curvature on soil erosion and deposition in hummock terrain. *Soil Science* 144, 209–217.
- [23] Roel, A., Plant, R., 2004. Factors underlying yield variability in two California rice fields. *Agron. J.* 96, 1481–1494.
- [24] Schabenberger, O., Pierce F.J., 2002. *Contemporary statistical models for the plant and soil sciences*. Taylor and Francis. CRC Press, Boca Raton, Florida.
- [25] Taylor, J.C., Wood, G.A., Earl, R., Godwin, R.J., 2003. Soil factors and their influence on within-yield crop variability. Part II: Spatial Analysis and Determination of Management Zones. *Biosystems Engineering*, 84, 441–53.

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.