

Análisis de componentes principales con datos georreferenciados

Una aplicación en agricultura de precisión

Principal component analysis with georeferenced data An application in precision agriculture

Mariano Córdoba ¹
Mónica Balzarini ¹

Cecilia Bruno ¹
José Luis Costa ²

Originales: Recepción: 18/08/2011 - Aceptación: 06/09/2011

RESUMEN

La utilización de nuevas tecnologías asociadas a la agricultura de precisión permite capturar información de múltiples variables en gran cantidad de sitios georreferenciados dentro de lotes en producción. Las covariaciones espaciales de las propiedades del suelo y el rendimiento del cultivo pueden evaluarse a través del análisis de componentes principales clásico (PCA). No obstante, como otros métodos multivariados descriptivos, el PCA no ha sido desarrollado explícitamente para datos espaciales. Nuevas versiones de análisis multivariado permiten contemplar la autocorrelación espacial entre datos de sitios vecinos. En este trabajo se aplican y comparan los resultados de dos técnicas multivariadas, el PCA y MULTISPATI-PCA. Este último incorpora la información espacial a través del cálculo del índice de Moran entre los datos de un sitio y el dato promedio de sus vecinos. Los resultados mostraron que utilizando MULTISPATI-PCA se detectaron correlaciones entre variables que no fueron detectadas con el PCA. Los mapas de variabilidad espacial construidos a partir de la primera componente de

ABSTRACT

New precision agriculture technologies allow collecting information from several variables at many georeferenced locations within crop fields. The spatial covariation of soil properties and crop yield data can be evaluated by principal component analysis (PCA). Nevertheless, PCA has not been explicitly developed for spatial data as other multivariate descriptive methods. Other multivariate techniques that include spatial autocorrelation among data of neighborhood sites have been recently developed. In this paper, we apply and compare two multivariate analyses, PCA and spatially constrained multivariate analysis methods (MULTISPATI-PCA). The latter incorporates the spatial information into multivariate analysis calculating Moran's index between the data at one location and the mean values of its neighbors. The results showed that MULTISPATI-PCA detected relations in the data that were not detected with PCA. The mapping of spatial variability from the first principal component was similar between PCA and MULTISPATI-PCA, but maps from the second component were different due to the variance correction by spatial autocorrelation. MULTISPATI-PCA method represents a crucial tool to map spatial variability within a field, and to identify homogeneous zones in a multivariate sense.

- 1 Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Estadística y Biometría. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba. Av. Valparaíso s/n. Ciudad Universitaria. C. C. 509. (5000) Córdoba. Argentina. marianoacba@agro.unc.edu.ar
- 2 Instituto Nacional de Tecnología Agropecuaria (INTA). Estación Experimental Balcarce. Ruta 226 Km 73,5. (7620) Balcarce. Buenos Aires. Argentina.

ambas técnicas fueron similares; no así los de la segunda componente debido a cambios en la estructura de co-variación identificada, al corregir la variabilidad por la autocorrelación espacial de los datos. El método MULTISPATI-PCA constituye una herramienta importante para el mapeo de la variabilidad espacial y la identificación de zonas homogéneas dentro de lotes.

Palabras clave

análisis multivariado • MULTISPATI-PCA
• PCA

Keywords

multivariate analysis • MULTISPATI-PCA
• PCA

INTRODUCCIÓN

En las últimas décadas se ha impulsado el desarrollo y la utilización de nuevas tecnologías para agricultura de precisión, que permiten capturar datos espaciales, *i. e.* aquellos asociados a una localización en el espacio. Los monitores de rendimiento acoplados a los equipos de cosecha y sensores de diferentes tipos, permiten recabar datos espaciales georreferenciados (establecidos sobre localizaciones de la superficie terrestre) produciendo cientos y miles de datos en áreas relativamente pequeñas y en un corto período (23).

Los sensores conectados a sistemas de posicionamiento global (GPS) están siendo utilizados, de manera más frecuente, para inferir distribución espacial de propiedades de los suelos y de los rendimientos, así como para delimitar zonas dentro del lote que luego pueden ser manejadas bajo agricultura sitio-específica (7).

Esta práctica agrícola implica la delimitación de subregiones dentro de los lotes, caracterizadas por expresar una combinación relativamente homogénea de factores de rendimiento, tales como textura del suelo, elevación, estado hídrico y nivel de nutrientes del suelo. La disponibilidad de datos para una gran cantidad de sitios dentro de un mismo lote, no sólo para variables de cultivo sino también para otras, como son las topográficas (elevación, profundidad de tosca) y las de suelo (conductividad eléctrica aparente), genera la necesidad de contar con técnicas de análisis de naturaleza multivariada. Tales técnicas deben además ser capaces de contemplar la característica espacial de los datos.

El análisis de las co-variaciones o correlaciones entre variables es un aspecto clave a tener en cuenta para comprender las interacciones que subyacen a los rendimientos. No obstante, es importante remarcar que la estructura de co-variación reflejada por un análisis multivariado clásico puede verse afectada por los patrones espaciales subyacentes en los datos.

La co-variación espacial de propiedades del sitio y el rendimiento de los cultivos ha sido estudiada a través de modelos geoestadísticos clásicos, que se basan en el concepto de autocorrelación para indicar la presencia de correlaciones o similitudes entre observaciones pertenecientes a sitios cercanos en el espacio, pero estos análisis se realizan variable por variable.

Otros estadísticos de autocorrelación espacial, como el índice de Moran (20) y el de Geary (16), también han sido usados para medir y analizar el grado de dependencia entre observaciones de una misma variable ubicadas en diferentes posiciones de un contexto geográfico. Estos análisis que se abordan variable a variable (univariados) dificultan la interpretación de la variabilidad conjunta, es decir, aquella causada por las relaciones entre variabilidad del rendimiento y de otras variables (6).

Una técnica multivariada comúnmente utilizada para el análisis de varias variables simultáneamente es el análisis de componentes principales (PCA) (2). Esta técnica permite identificar las variables que explican la mayor parte de la variabilidad total contenida en los datos, explorar las correlaciones entre variables y reducir la dimensión del análisis al combinar todas las variables en nuevos índices (variables sintéticas); cada una de estas nuevas variables constituye una componente principal (CP).

Los resultados del PCA se pueden visualizar en un gráfico denominado Biplot (14), el cual permite representar, en un plano óptimo para el estudio de variabilidad, las diferencias entre los sitios del lote y las variables que mejor explican las principales variaciones. La incorporación de la información geográfica puede realizarse *a posteriori* del PCA mediante la asignación de los valores de las componentes a cada uno de los sitios georreferenciados o ajustando semivariogramas (24) y usando otras herramientas de la geoestadística clásica (Kriging) para lograr mapas de variabilidad espacial por interpolación.

También se puede detectar la presencia de autocorrelación espacial de las CPs utilizando estadísticos de autocorrelación univariados como el índice de Moran o el de Geary *op. cit.* Dray *et al.* (10), proponen un método de análisis multivariado que incorpora la información espacial previo al análisis multivariado, el método es conocido como MULTISPATI-PCA. Éste se basa en el PCA pero incorpora la restricción dada por los datos espaciales mediante el cálculo del índice de Moran para medir la dependencia o correlación espacial entre las observaciones en un sitio y el promedio de las observaciones en el vecindario de ese sitio. Para delimitar los vecindarios utiliza una matriz de pesos espaciales determinando cuáles y cuántas observaciones cercanas a cada sitio conforman el vecindario.

Este análisis permite estudiar las relaciones entre las variables medidas (análisis co-variabilidad) y, al mismo tiempo, la estructura espacial (autocorrelación). La técnica ha resultado provechosa en estudios de ecología y suelos realizados a escala macrogeográfica (1,11).

Objetivo

- Ilustrar la aplicación del MULTISPATI-PCA con datos espaciales multivariados derivados del uso de maquinarias precisas en la pequeña escala (lote), comparando los resultados con los obtenidos vía un PCA clásico.

MATERIALES Y MÉTODOS

Datos

Se trabajó con datos provenientes de un lote en producción (65,4 ha) ubicado al sudeste bonaerense de la República Argentina, disponiendo de información de 672 sitios dentro del lote. Se compilaron valores georreferenciados de conductividad eléctrica aparente (CE) [mS m^{-1}] en dos profundidades 0-30 cm (CE30) y 0-90 cm (CE90), Elevación [m], profundidad de tosca (Tosca) [cm] y rendimiento de soja (RTO) [kg ha^{-1}].

Los valores de CE 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 (13). El instrumento fue calibrado, según las instrucciones del fabricante, antes de la recolección de los datos. Los datos de CE 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 del terreno también se midieron con un DGPS y se procesaron para obtener una precisión vertical de aproximadamente 3-5 cm. Las mediciones de profundidad de tosca se realizaron utilizando un penetrómetro hidráulico (Gidding) acoplado a un DGPS en una grilla regular de 30 m. Para cuantificar el rendimiento en grano del cultivo se utilizó un monitor de rendimiento acoplado a un equipo de cosecha conectados a un DGPS.

Los datos fueron sometidos a procedimientos de depuración para eliminar valores con alto error de medición. En esta instancia se incluyeron los datos que se encontraban entre la media ± 4 desvíos estándares. Debido a las diferentes resoluciones espaciales de las variables medidas, se promediaron los datos de CE30, CE90, Elevación y RTO, que se encontraban dentro de un radio de 14 m desde cada sitio donde se había registrado la variable Tosca. Se asignaron las medias de estas variables a cada uno de los sitios de medición de Tosca. Este procedimiento se realizó utilizando los softwares Idrisi Taiga (12) e InfoStat (9). La base de datos resultante estuvo conformada por $n = 672$ sitios (filas) y $p = 5$ variables (columnas).

Procedimientos de análisis

MULTISPATI-PCA

La base de datos provenientes de maquinarias precisas fue analizada utilizando MULTISPATI-PCA (10). Este método permite tener en cuenta la autocorrelación espacial de valores de la variable de interés en diferentes sitios de muestreo a través del análisis de vecinos de los sitios de muestreo.

Para ello, introduce una matriz de ponderación $\mathbf{W}_{n \times n}$ en el PCA de la matriz de datos originales ($n \times p$). Esta matriz es una representación matemática de la distribución geográfica de los sitios bajo estudio (3). Los pesos espaciales reflejan la ausencia ($w_{ij} = 0$) y la presencia o intensidad ($w_{ij} > 0$) de las relaciones espaciales entre los sitios dentro del área de estudio. La matriz de ponderación espacial puede

ser representada en forma gráfica (como gráficos de vecindarios), donde los nodos corresponden a los sitios del lote y los bordes a pesos espaciales no nulos. La especificación más simple de un vecindario es una matriz de conectividad \mathbf{C} , en la que $c_{ij} = 1$ si las unidades espaciales i y j son vecinos y $c_{ij} = 0$ en caso contrario.

Existen diferentes opciones o alternativas metodológicas para definir los vecindarios que dependen de los diferentes tipos de muestreo presente en los datos (grilla regular, irregular o transectas) (4). Para muestreos irregulares los métodos se basan en el gráfico de Gabriel (15), la triangulación de Delaunay (18), los vecinos más cercanos (8) y la distancia Euclídea entre otros. En este trabajo la red de vecindarios fue definida en función de la distancia Euclídea entre puntos vecinos contiguos.

Análisis de Componentes Principales

Encuentra los pesos o ponderaciones para cada variable con el fin de construir combinaciones lineales de variables capaces de maximizar la varianza entre los sitios de muestreo. Las combinaciones lineales obtenidas (CPs) son ortogonales (independientes) y en conjunto explican toda la variabilidad de los datos originales. La primera componente (CP1) explica la mayor parte de la variación total en el conjunto de datos y la segunda (CP2), la mayor parte de la variabilidad remanente o no explicada por la CP1.

Criterios de comparación

Los resultados obtenidos por PCA y MULTISPATI-PCA sobre los datos de agricultura de precisión, se compararon según distintos criterios; entre ellos la varianza explicada por las CPs. Para el PCA la varianza de cada CP es igual al autovalor asociado, mientras que para MULTISPATI-PCA, los autovalores asociados a las CPs son equivalentes a la varianza espacial (no varianza total) o varianza espacialmente estructurada. También se analizó la presencia de autocorrelación en las CPs con el índice de Moran. Se comparó la pérdida de la inercia (varianza espacial vs. varianza total) y el aumento de la información espacial (índice de Moran del MULTISPATI-PCA vs. índice de Moran del PCA).

La representación gráfica de las dos primeras CPs del MULTISPATI-PCA y el PCA se utilizó para observar la estructura de co-variación entre las variables (CE30, CE90, Elevación, RTO y Tosca). Para obtener mapas de variabilidad espacial multivariada, se usó predicción Kriging universal (24) sobre semivariogramas de la CP1 del PCA y la CP1 del MULTISPATI-PCA. Una medida cuantitativa de la similitud entre los valores de las dos primeras CPs de cada análisis se obtuvo mediante Procrustes Generalizado (17).

Para realizar los análisis estadísticos se utilizó el software libre R (www.r-project.org) mediante una interfase que lo conecta con el software InfoStat (9). La librería `ade4` (5) fue empleada para el PCA, Procrustes y MULTISPATI-PCA; para este último también se usó la librería `spdep` (4).

Finalmente para el análisis geoestadístico se utilizaron las librerías `gstat` (21) y `geoR` (22). Las sentencias correspondientes se pueden obtener solicitando su envío vía e-mail al primer autor.

RESULTADOS

En las tablas 1 y 2 se presentan las varianzas y los coeficientes de autocorrelación de cada una de las CPs generadas a partir de MULTISPATI-PCA y PCA, respectivamente. Como se establece en la literatura, MULTISPATI-PCA maximiza el producto entre la varianza espacial y la autocorrelación (10) mientras que PCA maximiza la varianza total (2). Los resultados muestran que con MULTISPATI-PCA se explica una menor proporción de la varianza acumulada en los dos primeros ejes o CPs, en esta aplicación, respecto de PCA (1,406 vs. 1,503 para el eje 1 y 0,945 vs. 1,065 para el eje 2).

No obstante, los valores del índice de Moran calculados para las dos primeras CPs sugieren que la estimación de autocorrelación aumentó cuando se usó MULTISPATI-PCA respecto de la contenida en las CPs del PCA (0,227 vs. 0,194 para el eje 1, 0,214 vs. 0,112 para el eje 2). Este resultado permitiría una visualización mejor de la variabilidad espacial. Por el contrario, a nivel de las CPs 3 y 5 este comportamiento fue inverso. Luego, se puede decir que no siempre la varianza disminuye a causa de la estimación de autocorrelación y que la maximización de la variabilidad espacial depende de las características de la autocorrelación presente.

Tabla 1. Autovalores, varianza espacial e índices de Moran de las componentes principales generados a partir de MULTISPATI-PCA (Dray *et al.*, 10) sobre cinco variables de suelo y rendimiento.

Table 1. Eigenvalues, spatial variance and Moran's indexes of principal components generated from MULTISPATI-PCA (Dray *et al.*, 10) on five soil variables and yield.

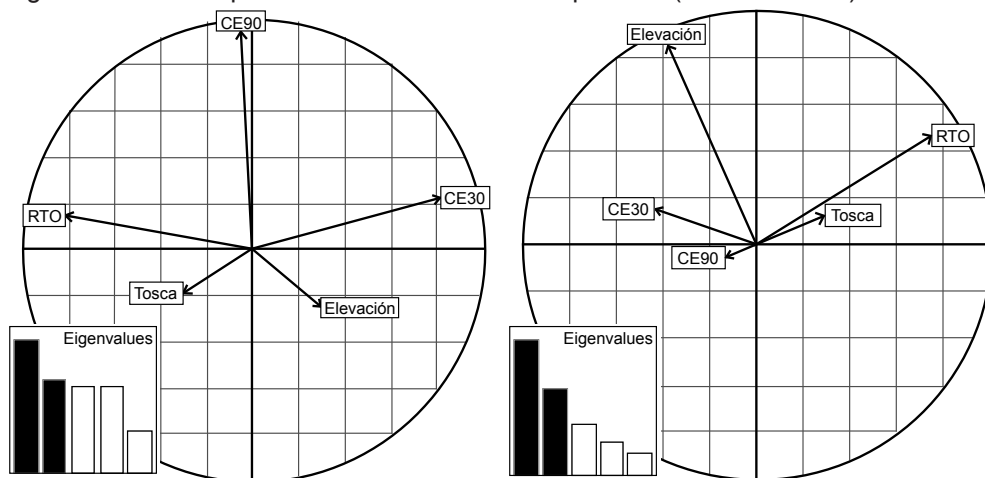
Autovalores	Varianza espacial	Proporción (varianza espacial)	Proporción acumulada	Índice de Moran
0,319	1,406	0,281	0,281	0,227
0,202	0,945	0,189	0,47	0,214
0,120	1,057	0,211	0,681	0,114
0,078	0,958	0,191	0,872	0,082
0,053	0,634	0,127	1,000	0,083

Tabla 2. Autovalores (varianza) e índices de Moran de las componentes principales generadas a partir del PCA sobre cinco variables de suelo y el rendimiento.

Table 2. Eigenvalues (variance) and Moran's indexes of principal components generated from PCA on five soil variables and yield.

Autovalores (varianza)	Proporción	Proporción acumulada	Índice de Moran
1,503	0,301	0,301	0,194
1,065	0,213	0,514	0,112
0,979	0,196	0,709	0,196
0,968	0,194	0,903	0,085
0,485	0,097	1,000	0,180

En la figura 1 se muestra la representación gráfica de las dos primeras componentes del PCA (izquierda) y también de las dos primeras componentes del MULTISPATI-PCA (derecha) así como los autovalores asociados a cada uno de esos ejes (gráfico de barras). Las barras de color negro corresponden a la cantidad de ejes seleccionados que fueron utilizados para la representación gráfica e interpretación de la variabilidad subyacente, en este caso las dos primeras CPs. Las barras negras junto con las blancas indican la cantidad de ejes que se obtuvieron en el análisis. La altura de cada barra representa la proporción de la variabilidad total reflejada por cada CP. Así, para MULTISPATI-PCA analizar las dos primeras CPs aporta suficiente información para el análisis, ya que las siguientes CP no representan una contribución importante (barras blancas).



El gráfico de barras de la esquina inferior izquierda representa los autovalores, las dos barras de color negro corresponden a los dos ejes que se utilizan para la representación gráfica y su altura determina la proporción de la variabilidad que explica cada componente.

The eigenvalues bar chart is drawn in the lower left corner, with the two black bars corresponding to the two axes used to draw the graphical representation and its height determines the proportion of the variability that explains each component.

Figura 1. Representación gráfica de los dos primeros ejes del PCA (izquierda) y MULTISPATI-PCA (derecha) que muestra la correlación entre las variables y entre éstas y las componentes principales.

Figure 1. Graphical representation of the two first axes of PCA (left) and MULTISPATI-PCA (right) and their eigenvalues. The correlation between variables and between these and the principal components is shown.

Mientras que para el PCA la selección del número de CP a tener en cuenta para la interpretación de la variabilidad es menos trivial ya que, la tercera y cuarta CP explican la misma proporción de variabilidad que la segunda CP; sin embargo en este trabajo, para fines prácticos, se interpretan sólo dos dimensiones. La variación en el eje 1, para ambos métodos, fue fuertemente impulsada por el RTO (variable de mayor proyección sobre el eje de las abscisas). No obstante, se visualizaron diferencias entre ambos análisis a nivel de la estructura de co-variación entre las dos mediciones de CE.

La CE90 se correlacionó con la CE30 cuando se realizó el análisis multivariado con restricción espacial (MULTISPATI-PCA) como sugiere la presencia de un ángulo pequeño entre los vectores de estas variables (figura 1 derecha, pág. 33).

Mientras que esta correlación, que es agronómicamente esperable ya que se trata de CE medida en el mismo sitio a dos profundidades, no fue rescatada por el PCA según muestra el ángulo casi recto entre los dos vectores de la variable (figura 1 izquierda, pág. 33).

Las relaciones entre Tosca y RTO y Tosca y Elevación no cambiaron de manera importante como para impactar las conclusiones como sí sucedió con las correlaciones entre CE30 y CE90.

El reposicionamiento de CE90 sobre el primer eje de MULTISPATI-PCA para indicar la presencia de correlación negativa entre RTO y CE, produjo un cambio de ponderación de las variables sobre la CP2 (eje de ordenadas) haciendo que ésta quede más correlacionada con la Elevación, permitiendo así analizar la variabilidad del rendimiento desde otra dimensión distinta a la de la CE.

La profundidad de Tosca estuvo en ambos análisis correlacionada positivamente con el RTO, señalando que aquellos lugares con mayor profundidad de Tosca o con menos impedimentos físicos, fueron los de mayor rendimiento.

Las figuras 2 y 3 (pág. 35) muestran los mapas logrados mediante la asignación a cada uno de los sitios georreferenciados de los valores de la CP1 y CP2 del PCA y MULTISPATI-PCA, respectivamente.

Para facilitar la visualización comparativa de ambos mapas se multiplicaron por -1 los valores de la CP1 del MULTISPATI-PCA, logrando así que valores mayores de la variable sintética correspondiesen a mayores valores de CE30 y menor RTO como sucede en la CP1 del PCA.

Los mapas de variabilidad espacial multivariada logrados con la CP1 de ambos métodos resultaron similares. Este resultado se justifica en el hecho de que en ambos análisis la CP1 se encuentra altamente correlacionada positivamente con CE30 y negativamente con RTO.

En estos mapas debe observarse que los sitios con valores mayores de la CP1 (cuadrados negros) corresponden a sectores del lote donde se obtuvo menor RTO en cualquiera de los dos mapas logrados (figura 2 derecha e izquierda, pág. 35).

Por otra parte, en los mapas construidos a partir de la CP2 del PCA (figura 3 izquierda, pág. 35) se representa principalmente la variación de la CE90 mientras que en el mapa logrado con la CP2 del MULTISPATI-PCA la variabilidad es debida principalmente a la Elevación (figura 3 derecha, pág. 35); los cuadrados negros representan aquí valores de mayor Elevación que los cuadrados blancos según pone de manifiesto la estructura de correlación de las variables (figura 1, pág. 33).

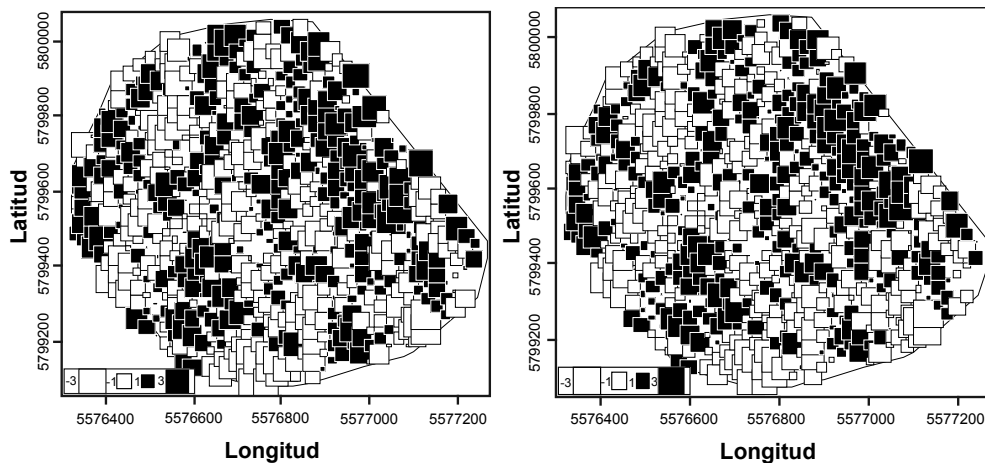


Figura 2. Mapas de variabilidad espacial multivariada logrados con la CP1 del PCA (izquierda) y CP1 del MULTISPATI-PCA (derecha).

Figure 2. Maps of multivariate spatial variability from the CP1 of PCA (left) and MULTISPATI-PCA (right).

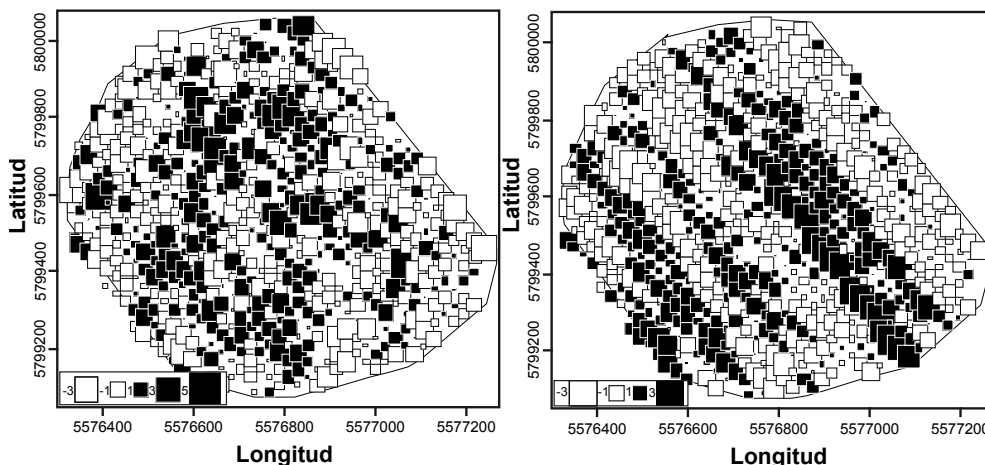


Figura 3. Mapas de variabilidad espacial multivariada logrados con la CP2 del PCA (izquierda) y CP2 del MULTISPATI-PCA (derecha).

Figure 3. Maps of multivariate spatial variability from the CP2 of PCA (left) and MULTISPATI-PCA (right).

En las figuras 4 y 5 (pág. 36) pueden observarse los mapas de variabilidad multivariada obtenidos como resultado del ajuste de semivariograma paramétrico y posterior interpolación por Kriging universal para las CP1 y CP2 del PCA y MULTISPATI-PCA, respectivamente. Este análisis permite realizar predicciones de la variable en los sitios no muestreados considerando la estructura de correlación espacial a través de los semivariogramas teóricos ajustados. Los patrones espaciales

son similares a los observados en las figuras 2 y 3 (pág. 35), pero se mejora la representación de la variabilidad espacial para la identificación de zonas homogéneas intralote, aspecto clave para la implementación del manejo sitio-específico.

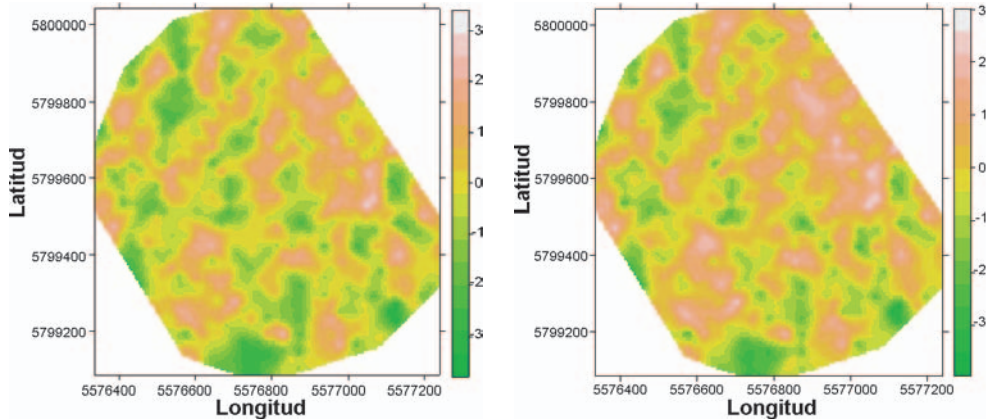


Figura 4. Mapas obtenidos por interpolación (Kriging) de la CP1 del PCA (izquierda) y CP1 del MULTISPATI-PCA (derecha).

Figure 4. Maps obtained by interpolation (Kriging) from the CP1 of PCA (left) and MULTISPATI-PCA (right).

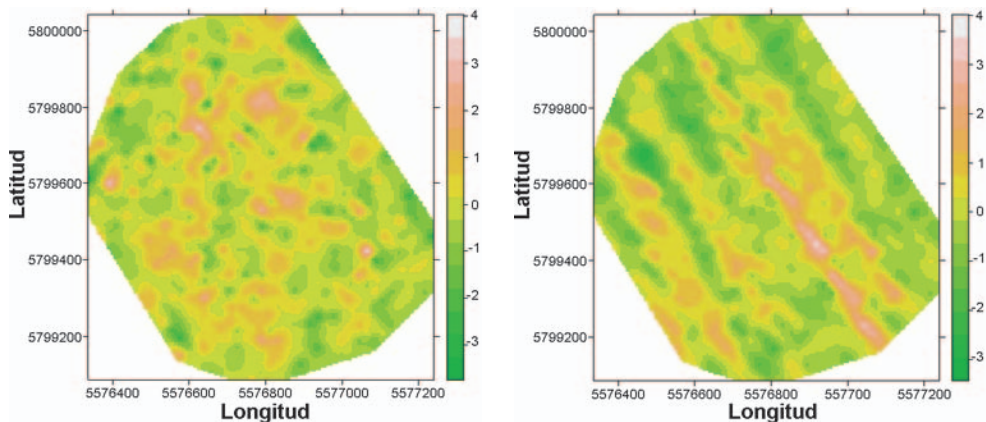


Figura 5. Mapas obtenidos por interpolación (Kriging) de la CP2 del PCA (izquierda) y CP2 del MULTISPATI-PCA (derecha).

Figure 5. Maps obtained by interpolation (Kriging) from the CP2 of PCA (left) and MULTISPATI-PCA (right).

Los resultados del análisis de Procrustes Generalizados aplicados sobre las configuraciones de sitios obtenidas a partir de las dos primeras CPs probaron una concordancia significativa ($p < 0,01$) y alta (correspondencia 0,69) entre los ordenamientos de los sitios vía las dos primeras CPs de los distintos análisis multivariados ilustrados (un valor de coeficiente de 1 indica una correspondencia perfecta entre las dos configuraciones).

DISCUSIÓN

La agricultura de precisión está demandando nuevas técnicas de análisis, entre las que se encuentran combinaciones de análisis geoestadísticos y multivariados para capturar la naturaleza de la observación espacial multivariada (1, 6, 11, 19, 24).

En este trabajo se aplicaron dos métodos para analizar datos espaciales multivariados (PCA y MULTISPATI-PCA) en una base de datos derivada del uso de maquinarias de agricultura de precisión, como son monitores de rendimientos acoplados a cosechadoras de granos y sensores de CE, medición de profundidad de tosca y elevación. Estas variables de suelo impactan altamente en los rendimientos de cultivos de granos (7), como la soja en la región pampeana argentina y por lo tanto podrían, junto con el rendimiento constituir indicadores sensibles a las diferencias en terreno para favorecer la delimitación de zonas homogéneas y el posterior manejo de precisión de la producción.

Utilizando MULTISPATI-PCA fue posible realizar un PCA que incorpore las coordenadas espaciales de los sitios monitoreados para considerar la posible presencia de autocorrelación espacial y ajustar las estimaciones de varianza del proceso monitoreado.

Con estos ajustes espaciales se detectó la co-variación entre CE30 y CE90 la cual no fue identificada por PCA. Desde un punto de vista agronómico se supone que es más probable que ambas variables (CE30 y CE90) estén correlacionadas a que no lo estén. Se trata de dos mediciones de una misma variable a dos profundidades diferentes pero en un mismo sitio. Otros autores han mostrado que estas variables pueden estar correlacionadas (7, 19). Aun cuando el nivel de concordancia entre las dos primeras CPs de ambos métodos puede ser significativo y alto, como se ha reportado en otras aplicaciones del PCA con restricción espacial respecto del PCA clásico (1, 11), es de esperar divergencia entre los ordenamientos de los puntos o sitios de muestreos.

Los resultados observados en la presente aplicación sostienen la idea de que los mapas de variabilidad espacial logrados por estas técnicas pueden ser diferentes en mayor o menor medida debido a que en MULTISPATI-PCA la varianza es corregida por la autocorrelación espacial. La incorporación de las coordenadas espaciales *a priori* del análisis hizo que la selección del número de CPs para la interpretación de la variabilidad fuese más clara que en PCA clásico, resultado también publicado por otros autores que trabajaron en escalas espaciales mayores o regionales y no a nivel intralote (11).

El método MULTISPATI-PCA constituye una herramienta estadística promisoría para la delimitación de zonas homogéneas en sentido multivariado, que podría ser utilizado en el mapeo de la variabilidad conjunta de variables de suelo y rendimiento capturadas dentro de lotes monitoreados con maquinarias precisas, tal como se ha realizado con otras técnicas multivariadas en bases de datos de agricultura de precisión (6, 19).

CONCLUSIONES

Los resultados mostraron que la incorporación de las coordenadas espaciales de los datos en el PCA permitió estimar la autocorrelación espacial en sentido multivariado y detectar relaciones subyacentes en el ensamblaje de variables originales, que si bien eran esperables agronómicamente, no fueron detectadas por el PCA cuando se ignoró la característica de datos espaciales de los vectores de variables asociados a los distintos sitios monitoreados por maquinarias precisas.

BIBLIOGRAFÍA

1. Arrouays, D.; Saby, N. P. A.; Thioulouse, J.; Jolivet, C.; Boulonne, L.; Ratié, C. 2011. Large trends in French topsoil characteristics are revealed by spatially constrained multivariate analysis. *Geoderma*, 161: 107-114.
2. Balzarini, M.; Teich, I.; Bruno, C.; Peña, A. 2011. Making genetic biodiversity measurable: a review of statistical multivariate methods to study variability at gene level. *Revista de la Facultad de Ciencias Agrarias. Universidad Nacional de Cuyo. Mendoza. Argentina.* 43(1): 261-275.
3. Bivand, R. 2008. Implementing representations of space in economic geography. *J. Reg. Sci.* 48(1): 1-27.
4. Bivand, R.; Altman, M.; Anselin, L.; Assunção, R.; Berke, O.; Bernat, A.; Blanchet, G.; Blankmeyer, E.; Carvalho, M.; Christensen, B.; Chun, Y.; Dormann, C.; Dray, S.; Halbersma, R.; Krainski, E.; Lewin-Koh, N.; Li, H.; Ma, J.; Millo, G.; Mueller, W.; Ono, H.; Peres-Neto, P.; Piras, G.; Reeder, M.; Tiefelsdorf, M.; Yu, D. 2010. *spdep: Spatial dependence: weighting schemes, statistics and models.* R package version 0.5-26. Disponible en: <http://CRAN.R-project.org/package=spdep>.
5. Chessel, D.; Dufour, A. B.; Thioulouse, J. 2004. The *ade4* package-I- One-table methods. *R News.* 4: 5-10.
6. Córdoba, M.; Bruno, C.; Costa, J.; Balzarini, M. 2011. Geoestadística multivariada en agricultura de precisión. 40° Jornadas Argentinas de Informática, 3° Congreso Argentino de Agrolinformática. Córdoba. p. 32-42.
7. 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. p. 139-165.
8. Cover, T. M.; Hart, P. E. 1967. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory.* 13: 21-27.
9. Di Rienzo, J. A.; Casanoves, F.; Balzarini, M.; Gonzalez, L.; Tablada, M.; Robledo, C. W. 2011. *InfoStat* versión 2011. Grupo. InfoStat. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba. Argentina. Disponible en: <http://www.infostat.com.ar>.
10. 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.
11. Dray, S.; Jombart, T. 2010. Revisiting Guerry's data: introducing spatial constraints in multivariate analysis. *Ann. Appl. Stat.*, in press.
12. Eastman, J. R. 2009. *IDRISI 16: The Taiga Edition* (Worcester, MA: Clark University).
13. Farahani, H. J.; Flynn, R. L. 2007. Map Quality & Zone Delineation as affected by Width of Parallel Swaths of Mobile Agricultural Sensors. *Precis. Agric.* 96: 151-159.
14. Gabriel, K. 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika.* 58(3): 453-467.
15. Gabriel, K. R.; Sokal, R. R. 1969. A New Statistical Approach to Geographic Variation Analysis. *Syst. Zool.* 18: 259-278.
16. Geary, R. 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician.* 5(3): 115-145.
17. Gower, J. C. 1971. A general coefficient of similarity and some properties. *Biometrics.* 27: 857-872.
18. Lee, D. T.; Schachter, B. J. 1980. Two Algorithms for Constructing a Delaunay Triangulation. *Int. J. Comput. Inform. Sci.* 9(3): 219-242.
19. 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 Till. Res.* 106: 335-343.
20. Moran, P. 1948. The interpretation of statistical maps. *J. Roy. Stat. Soc. B.* 10: 243-251.
21. Pebesma, E. J. 2004. Multivariable geostatistics in S: the *gstat* package. *Computers & Geosciences.* 30: 683-691.

22. Ribeiro Jr, P. J.; Diggle, P. J. 2001. geoR: a package for geostatistical analysis. R News. 1(2): 15-18.
23. Roel, A.; Terra, J. 2007. Muestreo de suelos y factores limitantes del rendimiento. Introducción a la agricultura de precisión. En: Agricultura de Precisión: Integrando conocimientos para una agricultura moderna y sustentable. Bongiovanni, R.; Mantovani, E. C.; Best, S.; Roel, A. (ed.), PROCISUR/IICA. Montevideo. Uruguay. p. 65-80.
24. Schabenberger, O.; Pierce, F. J. 2002. Contemporary statistical models for the plant and soil sciences. Taylor and Francis. N. Y. CRC Press, Boca Raton, Florida. 738 p.

Agradecimientos

A los productores, personal técnico y al Ing. Agr. Nahuel Peralta de la Estación Experimental Agropecuaria, 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 de la 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).