

Algoritmo para el análisis de la variabilidad temporal y espacial de suelos agrícolas a escala de lote

Mariano Córdoba¹, Cecilia Bruno¹, Mónica Balzarini¹

¹ Facultad de Ciencias Agropecuarias, CONICET, Universidad Nacional de Córdoba, Av. Valparaíso s/n cc 509, Córdoba-Argentina
mbalzari@agro.unc.edu.ar

Resumen. El análisis de datos georreferenciados en lotes agrícolas sirve para comprender la variabilidad espacial del suelo y realizar zonificaciones para agricultura sitio-específica. La variabilidad de estos datos a través del tiempo es menos analizada. En este trabajo se propone una metodología estadística para estimar tendencia temporal de variables edáficas y construir mapas de variabilidad espacio-temporal. El algoritmo se basa en el análisis de la tendencia interanual promedio de sitios, estimación de la varianza temporal por sitio para cada variable y delimitación de zonas homogéneas. Para ilustrar su aplicación, se analizó la tendencia interanual promedio y la variabilidad temporal de MO, P, pH y CE en 12 lotes de la región semiárida, abarcado una superficie de 2.240 ha bajo agricultura intensiva. Los resultados mostraron que la variación espacial en las características de suelo no es permanente, produciéndose significativos cambios en la delimitación de zonas homogéneas.

Palabras clave: Modelos Lineales Mixtos, correlación espacial, varianza temporal, mapas espacio-temporales.

1 Introducción

Los mapas de rendimiento y los mapas de suelo a escala de lote proveen indicadores de riesgos asociados con la producción y constituyen herramientas claves para implementar el manejo sitio-específico en agricultura de precisión. Para el manejo de la variabilidad espacial se usa el concepto de zonas homogéneas; éstas se interpretan como conjuntos de sitios con una misma combinación de factores limitantes de la producción y por tanto con una única demanda para un determinado input necesario para mejorar la producción [1]. Para la delimitación de zonas homogéneas se usan datos georreferenciados de rendimiento y/o de suelo-terreno.

Los primeros mapas de rendimiento construidos con este tipo de datos se usaron bajo el supuesto que el potencial productivo de las zonas delimitadas sería constante. La idea de que la variable medida en un determinado sitio del lote sería siempre relativamente menor/mayor que la del sitio de otra zona del mismo lote fue aún más fuerte para las características de suelo-terreno. La alta variabilidad de los rendimientos ocasionada por variaciones climáticas aleatorias, reforzó aún más la característica de “permanente” dada a las zonificaciones producidas a partir de variables de suelo. Sin embargo, aún en suelos manejados uniformemente, pueden existir considerables diferencias en los patrones espaciales de una misma variable

registrada en diferentes momentos de tiempo. Estos cambios temporales de la distribución espacial son más tratados en el análisis de datos de rendimientos que en el análisis de datos de suelo. La delimitación de zonas homogéneas suele realizarse con datos de rendimientos de más de un año [2], pero si se usan datos de suelo, generalmente éstos son de un año. Las variables de suelo medidas transversalmente (en un momento de tiempo) son utilizadas, conjuntamente con los rendimientos, para la delimitación de zonas homogéneas. La zonificación puede realizarse en estos casos de forma multivariada, es decir contemplando todas las variables simultáneamente [3]. Cuando los datos de suelo se obtienen por sitio en varios años, la zonificación podría depender no sólo de la variabilidad espacial sino también de la variabilidad temporal de las mediciones. La variabilidad temporal en las variables de suelo puede impactar las zonificaciones y consecuentemente las predicciones de rendimientos para años futuros. La tendencia promedio de los cambios de las variables medidas subsecuentemente en distintos años puede ser analizada con modelos estadísticos contemporáneos como los Modelos Lineales Mixtos (MLM) [4]. Estos modelos permiten contemplar las correlaciones esperadas entre datos obtenidos en distintos sitios del mismo lote. Lark y Stafford (1996) [5], Panetton y Barouillard (2001) [6], Blackmore et al. (2003) [2] y Marques da Silva (2006) [7] desarrollaron distintas metodologías para el análisis de la variabilidad temporal de datos georreferenciados. La combinación de modelos para datos longitudinales espaciales y técnicas de estimación de variabilidad temporal podría ser usada para construir zonificaciones que no solo consideren la variabilidad espacial de las variables sino también su estabilidad temporal. El objetivo de este trabajo fue proponer una metodología de análisis basada en MLM para estimar tendencia promedio y variabilidad temporal por sitio en variables de suelo usadas para la delimitación de zonas homogéneas.

2 Materiales y Métodos

Datos experimentales

Para ilustrar la metodología propuesta, se analizaron datos de un estudio llevado a cabo en un campo ubicado en el departamento Río Seco de la provincia Córdoba, Argentina (datos obtenidos por gentileza del Ing. Agr. Pablo Solfanelli). El área de relevamiento de datos pertenece a la región semiárida y se caracteriza por una topografía de sierra ondulada y una precipitación anual de 700 mm con veranos calurosos. Se analizó la tendencia temporal en un periodo de 12 años (con mediciones cada tres años: 2005, 2008 y 2011) de las variables MO (%), P (ppm), pH, y CE (dS/m). Los datos experimentales fueron recolectados en 12 lotes (post-desmonte) de 160 a 220 has cada uno, abarcado una superficie total de 2.240 ha bajo agricultura continua con rotaciones de cultivos anuales (Soja, Maíz y Trigo). Se obtuvieron entre 170 y 280 muestras de suelo en cada año; las muestras fueron georreferenciadas con un DGPS.

Procesamiento de datos

Los datos fueron pre-procesados para la eliminación de errores identificables mediante el uso de gráficos box-plot. Para cada variable se evaluó la tendencia

interanual promedio del área con un MLM con efecto fijo de año y de lote y estructura de correlación espacial para los términos de error aleatorio. El modelo con correlación espacial exponencial basada en el cálculo de distancias entre los sitios de muestreo georreferenciados fue comparado con el modelo de errores independientes, mediante la prueba de cociente de verosimilitud (LRT) basada en los estimadores REML (máxima verosimilitud restringida) de los parámetros de varianza y covarianza [8]. Ajustado el modelo de correlación espacial, se evaluó la reducción del modelo (eliminación de efecto lote) mediante prueba LRT basada en estimadores ML (máxima verosimilitud). Las diferencias encontradas entre año fueron evaluadas con la prueba LSD de Fisher ($\alpha=0.05$) para las medias ajustadas por la estructura de varianza y covarianza subyacente. Las diferencias encontradas pueden ser atribuidas consistentemente a causas climáticas o a causas reconocidas *a priori*, como podría ser la evolución del suelo post-desmonte sometido a agricultura. Para calcular la variabilidad temporal de cada variable, en un determinado sitio o punto muestral se usó la siguiente expresión sobre los valores predichos por el modelo ajustado, para cada sitio:

$$\delta_i^2 = \frac{\sum_{t=05}^{t=11} (Y_{t,i} - \bar{Y}_t)^2}{3} \quad (1)$$

donde δ_i^2 es la varianza temporal del sitio i ; t es el año (2005, 2008 y 2011); $Y_{t,i}$ es el valor de la variable de suelo predicho por el modelo para el año t en el sitio i ; \bar{Y}_t es la media de la variable para todo el lote en el año t .

La ecuación refleja la desviación del dato de cada sitio con respecto a la media temporal de la variable. Por definición, la varianza temporal presenta un valor bajo, si un determinado sitio del lote presenta un valor que está siempre cerca de la media general de la variable [2]. El dato de desviación estándar temporal, obtenido con la función de desviación estándar modificada descrita en (1), fue representado gráficamente en función de la media temporal de la variable en el sitio, usando una grilla de 60×60m. La media de la variable de suelo a través de los años fue trazada como referencia vertical del gráfico para identificar los sitios que se ubican por encima de la media. Mientras que el tercer cuartil de la desviación estándar temporal, denotado como P(75), fue trazado como referencia horizontal del gráfico con el objeto de identificar los sitios más inestables *i.e.* DE > P(75).

Usando los datos espaciales de las cuatro variables de suelo se delimitaron dos zonas homogéneas en cada año mediante análisis de cluster sobre variables sintéticas derivadas de un análisis de componentes principales con restricción espacial [3]. Cada zona fue particionada nuevamente según la variabilidad temporal de cada variable. De esta manera se obtuvieron mapas espacio-temporales delimitando cuatro clases de sitios: ZM1-DE > P(75), ZM1-DE < P(75), ZM2-DE > P(75) y ZM2-DE < P(75). Todos los análisis fueron realizados con el software InfoStat a través de su interface con el software R [9].

3 Resultados y Discusión

El modelo de correlación espacial ajustó mejor que el modelo de errores independientes para todas las variables (LRT, $p < 0.05$). El efecto de lote fue también significativo para MO y P (LRT $p < 0.05$). Las distribuciones de valores de todas las variables, excepto CE, fueron relativamente simétricas, con mayores cambios en la media que en la varianza a través de los años (Fig. 1).

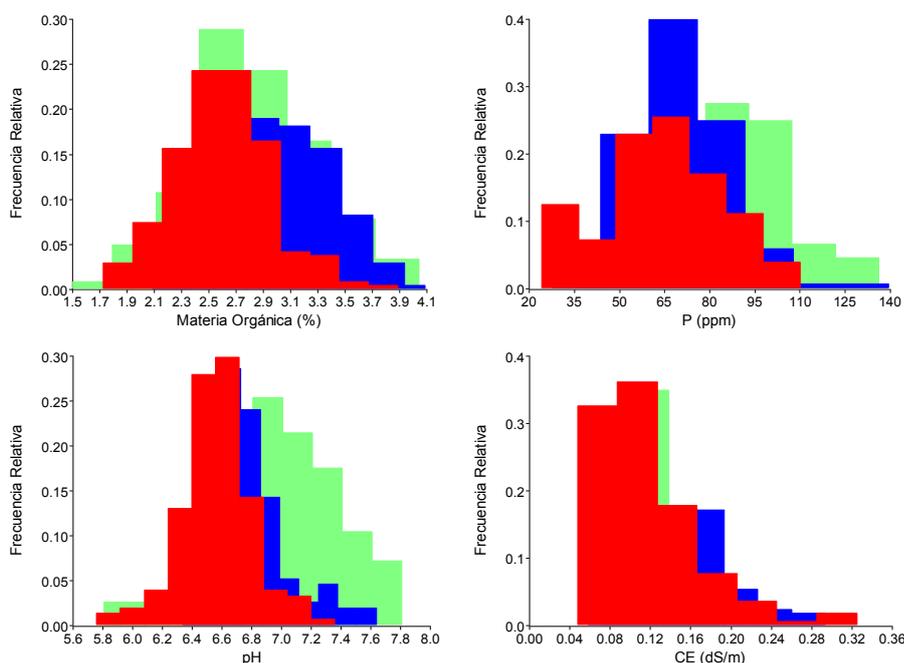


Fig. 1. Distribuciones de frecuencias de datos de MO, P, pH y CE en tres años (2005 (verde), 2008 (azul), y 2011 (rojo)).

La comparación de los valores promedios, de cada variable, entre los diferentes años de estudio (Tabla 1), indicó decrecimientos significativos y progresivos en P y pH (Fig. 2). La caída de la MO fue significativa en el noveno año ($P < 0.05$), mientras que los valores de CE permanecieron sin cambios temporal significativos y fueron los de mayor variabilidad espacial dentro de cada año. Estos cambios podrían indicar sitios con mayor/menor riesgo de obtención de rendimientos pobres, debido a correlaciones detectadas entre la CE y el rendimiento [3].

Tabla 1. Media, coeficiente de variación (CV), valores mínimos (Min.) y máximos (Max.) de variables de suelo.

Variable	Año	Media	CV	Min.	Max.
MO	2005	2.88	17.50	1.47	4.69
	2008	2.98	14.42	2.00	4.40
	2011	2.61	13.69	1.73	3.89
P	2005	85.54	24.15	35.54	136.21
	2008	70.13	22.56	27.83	139.95
	2011	64.02	32.79	24.30	115.65
pH	2005	7.06	5.51	5.81	7.81
	2008	6.80	3.64	6.30	7.74
	2011	6.57	3.58	5.76	7.36
CE	2005	0.13	15.22	0.07	0.19
	2008	0.13	40.77	0.05	0.32
	2011	0.12	41.19	0.04	0.33

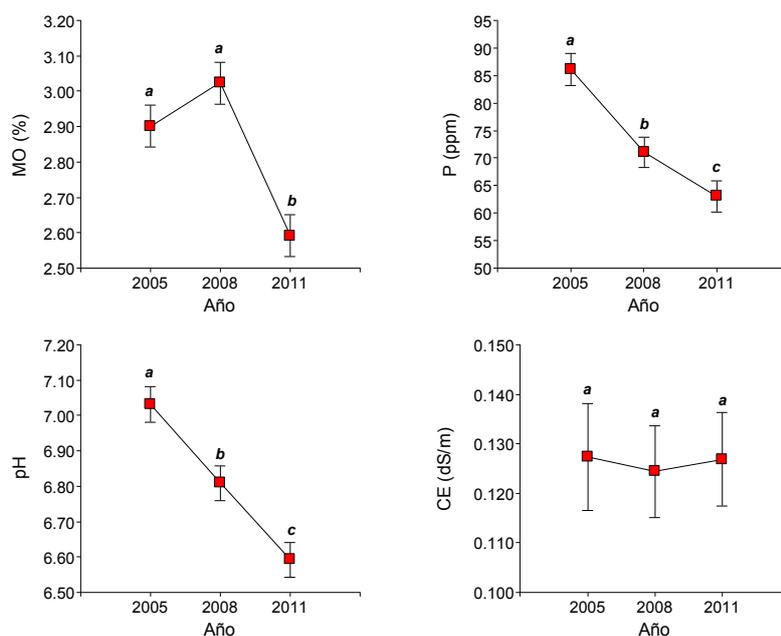


Fig. 2. Tendencia temporal en promedios de MO, P, pH y CE. Letras diferentes indican diferencias estadísticamente significativas ($p < 0.05$).

Los gráficos de dispersión de la variabilidad temporal en función de la media temporal de cada variable (Fig. 3), muestran que para MO, los sitios más inestables fueron los que tenían valores más alejado del promedio temporal (2.83%) y que la DE temporal de estos sitios fue relativamente alta (muy por encima del P(75)). Para CE, los sitios más inestables fueron los de valores mayores a la media (0.126 dS/m).

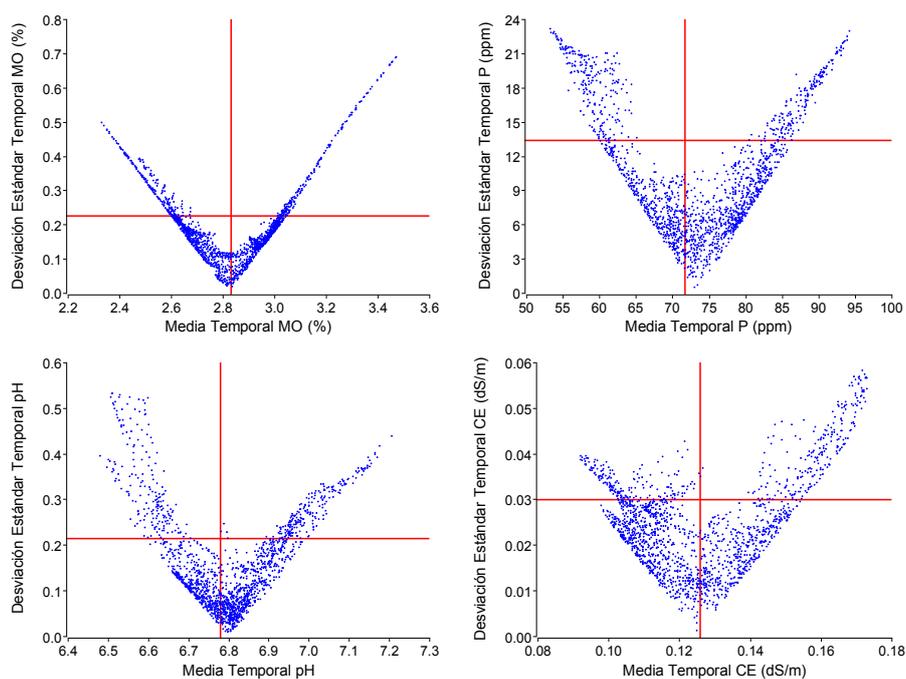


Fig. 3. Gráficos de dispersión de la variabilidad temporal en función de la media temporal de cada variable de suelo.

Los mapas de la Figura 4, muestran las zonificaciones logradas con los datos de las cuatro variables en cada año y su partición en relación a la inestabilidad temporal de cada variable. En el análisis de la variabilidad espacial, se delimitaron dos zonas homogéneas (ZM) cada año. La ZM 1 se caracterizó por presentar mayores contenidos de MO, P y CE y menor pH. Esta ZM mostró mayor proporción de sitios clasificados como inestables.



Fig. 4. Zonas de manejo para cada año obtenidas con datos de cuatro variables de suelo y particionadas en función de la inestabilidad temporal de cada variable. Sitios con desviación estándar temporal por encima del tercer cuartil de la desviación temporal (P(75)) son más inestables.

4 Conclusión

Los resultados obtenidos con el algoritmo propuesto mostraron que la variación espacial en las características de suelo no fue permanente, produciéndose significativos cambios en la delimitación de zonas homogéneas a través de los años.

Agradecimientos. Al Ing. Agr. Pablo Solfanelli y a la empresa agropecuaria que facilitó las determinaciones de suelo para esta investigación. A la Universidad Nacional de Córdoba y al CONICET por subsidiar desarrollos en el área del análisis Estadístico y el desarrollo informático para análisis de datos.

Referencias

1. Vrindts, E., Mouazen, A. M., Reyniers, M., Martens, K., Maleki, M. R., Ramon, H. and De Baerdemaeker, J.: Management zones based on correlation between soil compaction, yield and crop data. *Biosyst. Eng.*, 92, 419--428 (2005).
2. Blackmore, B.S., Godwin, R.J., Fountas, S.: The analysis of spatial and temporal trends in yield map data over six years. *Biosyst. Eng.*, 84(4), 455--466 (2003).
3. Córdoba, M., Balzarini, M., Bruno, C., Costa, J.: Identificación de zonas de manejo sitio-específico a partir de la combinación de variables de suelo. *Revista Corpoica*. 13,47--54. (2012).
4. Schabenberger, O., Pierce, F.J.: Contemporary statistical models for the plant and soil sciences. Taylor and Francis. CRC Press, Boca Raton, Florida (2002).
5. Lark, R., Stafford, J.: Classification as a first step in the interpretation of temporal and spatial variation of crop yield. *Ann. Appl. Biol.* 130, 111--121 (1997).
6. Panneton, B., Brouillard, M., Piekutowski, T.: Integration of yield data from several years into a single map. In: Third European Conference on Precision Agriculture , pp. 73--78. Agro Montpellier, Montpellier (2001).
7. Marques Da Silva, J.R.: Analysis of the Spatial and Temporal Variability of Irrigated Maize Yield. *Biosyst. Eng.*, 94, 337--349 (2006).
8. Pinheiro, J.C., Bates, D.M.: Mixed-Effects Models in S and S-PLUS. Springer, New York. (2004).
9. Di Rienzo, J.A., Casanoves, F., Balzarini, M.G., Gonzalez, L., Tablada, M., Robledo, C.W.: InfoStat versión 2013. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. URL <http://www.infostat.com.ar>.