

ZoneCompare: software para el análisis de ensayos comparativos de rendimientos sin repeticiones en agricultura de precisión

Pablo Paccioretti^{1,2}, Franca Giannini Kurina^{1,2}, Mariano Córdoba^{1,2},
Mónica Balzarini^{1,2}

¹Universidad Nacional de Córdoba, Facultad de Ciencias Agropecuarias, Cátedra de Estadística y Biometría. Córdoba, Argentina.

²Unidad de Fitopatología y Modelización Agrícola (UFyMA), INTA – CONICET.

Resumen. La agricultura de precisión facilita la instalación de ensayos en campos de productores. Es común la comparación de tratamientos agronómicos sobre macroparcels donde la cosecha se realiza con monitores de rendimiento. Cada dato tiene asociada su posición en el terreno (georreferencia), haciendo posible la identificación de modelos de variabilidad y correlación espacial de los rendimientos dentro del lote. Sin embargo, la automatización del proceso de captura de datos puede registrar valores raros, tanto globales como espaciales, que impactan sobre la modelización de la variabilidad espacial. En muchos ensayos en campos de productores se trabaja con unidades experimentales grandes, pero con una repetición por tratamiento. Consecuentemente, el análisis para la comparación de medias de tratamiento no puede ser realizado utilizando técnicas estadísticas clásicas. En este trabajo se presenta una aplicación web interactiva, que fue desarrollada utilizando funciones especializadas del lenguaje de programación R para atender las necesidades de procesamiento de datos de este tipo de ensayos. El software permite la depuración de datos de rendimiento dentro de cada parcela y la comparación de medias de tratamientos usando estimadores de varianza derivados de la información intralote que no requieren repeticiones.

1 Introducción

En las últimas décadas la realización de ensayos agronómicos con la finalidad de comparar rendimientos entre tratamientos utilizando tecnologías de precisión se ha incrementado. Estos ensayos son llevados a cabo tanto en lotes experimentales como en campos de productores. Atender a los principios básicos del diseño de experimento clásico (aleatorización, replicación, bloqueo) no siempre es posible. Algunos ensayos son diseñados sin repeticiones de cada tratamiento, es decir una parcela por tratamiento. Por lo tanto, los métodos de análisis clásicos como el análisis de la varianza no pueden ser empleados dado que no se puede estimar una varianza residual.

Los ensayos suelen realizarse dividiendo un lote, o una porción de él, con tantas parcelas como tratamientos se quieran evaluar y luego los rendimientos son registrados con monitores precisos en donde a cada dato se le asocia su posición en el terreno,

obteniendo numerosos datos para cada parcela. Al ser un proceso automático, se suelen registrar muchos datos erróneos [1]. La vasta cantidad de datos por parcela habilita la estimación de modelos geoestadísticos de variabilidad espacial intralote los cuales también pueden proveer estimadores de la variabilidad residual. Luego la estimación de estos modelos facilita la comparación de medias de tratamientos aún en situaciones donde estos no están repetidos. Sin embargo, el ajuste de modelos geoestadísticos puede verse afectado por la presencia de valores raros [2]. Luego, el procesamiento y análisis de estos ensayos con muchos datos, pero sin repeticiones demanda de herramientas de software capaces no solo de realizar la comparación estadística de los tratamientos sino también de preprocesar adecuadamente los datos de los monitores de rendimiento.

En este trabajo se describe una nueva aplicación de software, la cual ha sido desarrollada para automatizar, en una interfase simple y fácil de usar, un protocolo completo de análisis estadísticos para datos de monitores de rendimiento que incluye el preprocesado de la base de datos y la comparación estadística de rendimientos medios entre parcelas tratadas de diferente forma sin demandar repeticiones de tratamientos. Las facilidades del software para importar, limpiar, analizar y visualizar datos espaciales, en un único ambiente de computación, permite un entorno fácil para comparar medias de tratamientos en este tipo de ensayos.

2 Materiales y métodos

2.1 Software

El software, denominado ZoneCompare, fue desarrollado utilizando paquetes especializados del lenguaje de programación R [3], los cuales posibilitan la aplicación de protocolos de depuración univariada en el contexto de datos correlacionados espacialmente (Vega et al., 2019). La interfaz gráfica de usuario está generada principalmente con funciones del paquete *shiny* [5]. El manejo y depuración de datos espaciales se realiza con funciones del paquete *spdep* [6]. El software está alojado en un servidor de la Universidad Nacional de Córdoba (UNC). El código de programación se encuentra en <https://github.com/PPaccioretti/ZoneCompare>. Para comenzar una sesión de ZoneCompare, es necesario contar con conexión a internet se debe ingresar a <https://fastmapping.psi.unc.edu.ar/ZoneCompare>.

2.2 Datos de ilustración

Para ilustrar el uso del software se utilizó una base de datos proveniente de un ensayo comparativo de rendimiento realizado en un lote comercial (100 ha) sembrado con un cultivo de girasol (*Helianthus annuus* L.). El lote fue dividido en tres sectores, a dos de ellos se le aplicaron nuevos tratamientos de fertilización, mientras que una división o parcela fue manejada con la práctica usual del productor (tratamiento control). El ensayo fue cosechado en su totalidad utilizando cosechadora con monitor de rendimiento.

2.3 Propuesta de análisis embebida en ZoneCompare

El software automáticamente transforma el sistema de coordenadas geográficas a cartesianas Universal Transversal de Mercator (UTM). De esta manera las interpretaciones de distancias pueden realizarse en metros en vez de grados. En la depuración de datos, se remueven datos relacionados a los efectos bordes. Además, se eliminan valores que se encuentran por fuera del patrón general de los datos. Para esto, el conjunto de datos se limita dentro de valores mínimos y máximos esperables para la distribución de los rendimientos. Luego, calcula la media y la desviación estándar (DE), eliminando valores que se encuentren por fuera del intervalo media ± 3 DE. Teóricamente aproximadamente el 90% de los datos de una variable aleatoria se encuentra dentro de este rango. Los valores eliminados se contabilizan como outliers global. Por último, se eliminaron aquellos valores que, si bien su observación es posible para la distribución de la variable, difieren significativamente de su vecindario (outliers espacial). Para detectar estos valores, se utiliza el índice de autocorrelación espacial local de Moran (IM) y el gráfico de dispersión de Moran [7]. Si al menos uno de estos criterios detecta la observación como rara, ésta es eliminada automáticamente. Para calcular la autocorrelación espacial con el índice de Moran, es necesario generar una red de vecindarios, en este se consideraron sitios vecinos a aquellos que se encuentran dentro de una distancia de referencia (en el ejemplo, 35m).

El software provee de manera automática la comparación de medias entre tratamientos para el rendimiento. La comparación de medias se realiza con varianzas calculadas después del ajuste por correlación espacial según la siguiente expresión

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

donde Y_{ZonaX} es la media de la Zona X, Y_{ZonaW} es la media de la Zona W y $\tilde{\sigma}_{krig}$ es la mediana de la desviación estándar kriging [8].

3 Resultados

Para el ejemplo de ilustración, durante el proceso de depuración se removió el 48% de las observaciones por considerarlas como borduras o outliers, principalmente outliers espaciales. La comparación de medias arrojó diferencia significativa entre el tratamiento A y B, mientras que el tratamiento C (Testigo) no se diferenció de los demás tratamientos ($\alpha=0,05$). El output producido por el software se presenta en la Fig 1. El mismo contiene una tabla describiendo los motivos de eliminación de datos de monitores de rendimiento previo al análisis estadístico donde se especifica el porcentaje de distintos tipos de datos depurados dentro de cada una de las parcelas de análisis y la gráfica de comparación de medias.

Aplicado para analizar datos de monitores de rendimiento, el software permite la comparación de tratamientos o de zonas de manejo delimitadas por el ojo de un experto. Asimismo, puede utilizarse a escala regional para la comparación de medias entre regiones delimitadas previamente que son únicas o no se encuentran repetidas en el territorio de estudio.

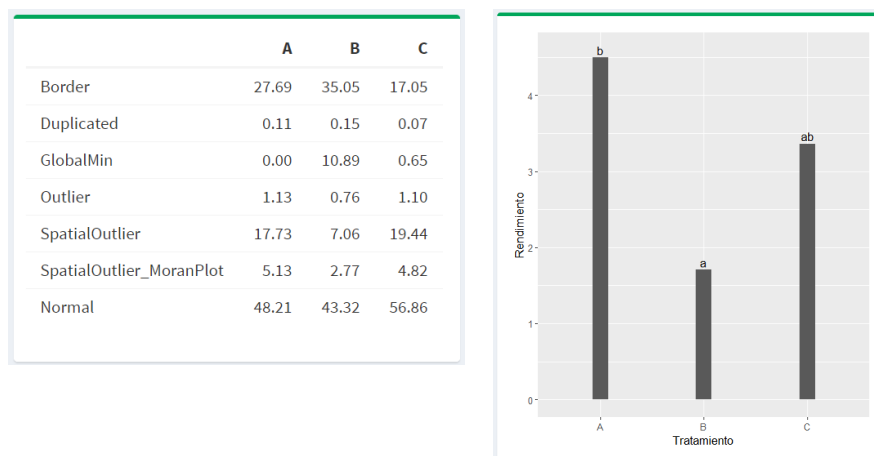


Fig. 1. Captura de pantalla de la información suministrada en el *output* de ZoneCompare.

4 Conclusión

La innovación metodológica del software ZoneCompare es la integración de una etapa de pre-procesamiento de datos orientada a eliminar anomalías en datos espaciales y la comparación de medias de tratamientos en contextos de ensayos sin repeticiones. La disponibilidad de aplicaciones basadas en R, con una interfaz amigable, permite mejorar las bases metodológicas y computacionales de aplicaciones para la agricultura de precisión.

4 Bibliografía

1. Sudduth, K.A., Drummond, S.T.: Yield Editor: Software for Removing Errors from Crop Yield Maps. *Agron. J.* 99, 1471–1482 (2007).
2. Vega, A., Córdoba, M., Castro-Franco, M., Balzarini, M.: Protocol for automating error removal from yield maps. *Precis. Agric.* 1–15 (2019).
3. R Core Team: R: A Language and Environment for Statistical Computing, <https://www.r-project.org/>, (2019).
4. Córdoba, M., Bruno, C., Costa, J.L., Balzarini, M.: Subfield management class delineation using cluster analysis from spatial principal components of soil variables. *Comput. Electron. Agric.* 97, 6–14 (2013).
5. Chang, W., Cheng, J., Allaire, J.J., Xie, Y., McPherson, J.: shiny: Web Application Framework for R, <https://cran.r-project.org/package=shiny>, (2019).
6. Bivand, R.S., Wong, D.W.S.: Comparing implementations of global and local indicators of spatial association. *TEST.* 27, 716–748 (2018).
7. Anselin, L.: The Moran scatterplot as an ESDA tool to assess local instability in spatial association. En: *Spatial Analytical.* p. 121. CRC Press (1996).
8. Taylor, J.A., McBratney, A.B., Whelan, B.M.: Establishing Management Classes for Broadacre Agricultural Production. *Agron. J.* 99, 1366:1376 (2007).