



Universidad
Nacional
de Córdoba



FCA
Facultad de Ciencias
Agropecuarias

Metodos cuantitativos para la investigación agropecuaria

“Comparación de métodos de estimación de semivariogramas para el mapeo de variabilidad espacial del rendimiento intralote”

Autores:

- *Barberis, Gaston*
- *Campodonico, Andres*
- *Cesaratto, Pedro*
- *Di Noto, Guillermo*
- *Marinzalda, Juan Manuel*

Tutor:

- *Ing. Agrónomo Córdoba, Mariano*



Contenido

TITULO	1
ÍNDICE DE TABLAS Y FIGURAS	3
AGRADECIMIENTOS	4
PALABRAS CLAVES	5
INTRODUCCIÓN	7
MATERIALES Y MÉTODOS.....	9
DATOS.....	9
MÉTODOS	9
RESULTADOS Y DISCUSIÓN.....	11
CONCLUSIÓN	16
CONSIDERACIONES FINALES	16
BIBLIOGRAFÍA	17

ÍNDICE DE TABLAS Y FIGURAS

Figura 1 Semivariograma. Se representan los 3 parámetros que lo definen: Rango (R), Sill Parcial (C) y Nugget (C_0).

Tabla 1. Medias resumen. Media, Mínimo y Máximo del rendimiento promedio y CV de 20 mapas de rendimiento y diferencias entre Medias de rendimiento y CV obtenidos mediante la prueba Wilcoxon de datos depurados y no depurados con sus respectivos p-valor.

Tabla 2. Resultados de las diferencias entre las Medias de los parámetros de los semivariogramas y del grado de estructuración espacial (RSV) con sus respectivos p-valor obtenidos mediante Wilcoxon para datos depurados con estimador de Matheron y datos no depurados con estimador de Cressie.

Figura 3. Semivariogramas empíricos y teóricos obtenidos a partir del estimador robusto de Cressie para datos sin depurar y el estimador de Matheron para datos depurados.

Figura 4. Mapas de variabilidad espacial de rendimiento obtenidos a partir de parámetros de semivariogramas estimados a partir del estimador robusto de Cressie para datos sin depurar y el estimador de Matheron para datos depurados.

AGRADECIMIENTOS

Agradecer a nuestro tutor el Ing. Agrónomo Mariano Córdoba tanto por el aporte de datos de importancia como por su actividad de coordinación y predisposición para el desarrollo del presente trabajo.

Por otra parte resaltar el aporte del Ing. Agrónomo Alejandro Tobal al presente trabajo, cuya participación facilitó la integración ética al mismo.

RESUMEN

En los últimos años se ha dado un gran impulso al desarrollo e implementación de nuevas tecnologías en el área de la agricultura, entre estas los monitores de rendimiento son las de mayor uso por parte de los productores. Estos permiten recolectar datos de rendimiento asociados a una localización en el espacio para diferentes sitios y generar mapas de rendimiento utilizando métodos de interpolación espacial. En el presente trabajo comparan dos alternativas para caracterizar la estructura de correlación espacial del rendimiento en lotes agrícolas. La primera alternativa utiliza datos depurados y realiza la estimación del semivariograma mediante el estimador de Matheron. La segunda aproximación, realiza el cálculo del semivariograma sobre datos sin depurar, pero utilizando estimador robusto de Cressie. Para la comparación se utilizan datos de 20 mapas de rendimiento de los principales cultivos de grano de la región pampeana Argentina. Los resultados mostraron que existe diferencias importantes en los parámetros estimados de los emivariogramas, pero la relación entre estos y los patrones de variación fueron similares, tanto utilizando el estimador clásico para datos depurados como el estimador robusto para datos sin depurar.

PALABRAS CLAVES

Agricultura de precisión - kriging - autocorrelacion espacial.

INTRODUCCIÓN

La variabilidad espacial en las propiedades o factores determinantes de la producción en sistemas agrícolas, ha sido observada por los productores desde los inicios de la agricultura (Bullock et al., 2007). Así, la variación espacial de los rendimientos es el resultado de una compleja interacción de factores biológicos, edáficos, topográficos, antropogénicos o de manejo y climáticos ¹(Bongiovanni *et al.*, 2006).

En los últimos años se ha dado un gran impulso al desarrollo e implementación de nuevas tecnologías en el área de la agricultura², estas permiten recoger distintos tipos de datos espaciales³, i.e. datos de diferentes variables asociados a una localización en el espacio para diferentes sitios⁴ (Córdoba et al., 2015). En particular, la tecnología de agricultura de precisión (AP) comienza a dar sus primeros pasos en Argentina alrededor del año 1996, en la Estación Experimental Agropecuaria INTA ⁵ Manfredi, cumpliendo este organismo un importante rol en la difusión⁶, experimentación y capacitación⁷ en el uso de dichas tecnologías.

La AP⁸ y la tecnología de gestión de manejo por ambientes^{9,10} han tenido una importante difusión en los últimos años, haciendo que la mayoría de los actores del sector agrícola hayan adoptado al menos algunas de estas¹¹. El óptimo uso del gran volumen de datos derivado de maquinarias de AP y de sistemas de sensores, depende fuertemente de las capacidades para explorar los datos espaciales que de ellos se obtienen (Bullock y Lowenberg-DeBoer, 2007). La covariación espacial de los rendimientos de los cultivos puede ser estudiada a través de una gran diversidad de métodos y modelos estadísticos, los más usados son los modelos geoestadísticos clásicos (Cressie, 1993; Schabenberger y Gotway, 2004; Webster y Oliver, 2007). En geoestadística uno de los objetivos es cuantificar la estructura espacial de una variable georreferenciada mediante el uso de semivariogramas y utilizar sus parámetros para realizar predicciones con los métodos de interpolación como kriging. Utilizando estos métodos se obtienen mapas que permiten visualizar los patrones de variación en el espacio de la variable analizada.

¹ Indicador 44: Uso Sustentable de la biodiversidad-restauración hábitats naturales

² Indicador 43: Uso sustentable de recurso Energía.

³ Indicador 2: Propuesta de valor.

⁴ Indicador 1: Estrategias para la sustentabilidad.

⁵ Indicador 14: Participación en el desarrollo de políticas públicas. Público de interés.

⁶ Indicador 7: Participación de los públicos de interés.

⁷ Indicador 27: Compromiso con el desarrollo personal.

⁸ Indicador 11: Competencia leal.

⁹ Indicador 37: Enfoque y gestión de acciones relacionadas con el cambio climático.

¹⁰ Indicador 40: Prevención de la polución.

¹¹ Indicador 12: Prácticas anti corrupción.

Los datos del monitor de rendimiento comúnmente contienen valores erróneos debidos a diferentes fuentes, entre ellas los errores de registro por altura de cabezal, retardos en el inicio y fin de la cosecha, valores de rendimientos fuera de rango biológico esperado, pequeños parches o franjas angostas con rendimientos extremadamente altos o bajos que no están relacionados con sus vecinos inmediatos, pequeños segmentos y puntos de rendimientos superpuestos (grabados con iguales coordenadas). La remoción de estos datos es importante para evitar estimaciones erróneas de semivariogramas y en las estimaciones de los rendimientos, como así también para mejorar las comparaciones con otras capas de información, y consecuentemente, mejorar la precisión de las decisiones de manejo.

Uno de los estimadores de semivariograma más utilizado es el método de los momentos, que recibe el nombre de estimador clásico de semivariograma propuesto por Matheron (1962). Sin embargo, otros autores, como Genton (1998) y Mingoti y Rosa (2008), señalan que este estimador se ve muy afectado por la presencia de valores atípicos en los datos. Los outliers, o valores atípicos, son observaciones con valores que se encuentran fuera del patrón general o distribución del conjunto de datos. La eliminación de los outliers, previo al análisis, es fundamental para garantizar que las decisiones tomadas a partir del análisis sean las correctas. Sin embargo, al descartar los outliers se eliminan los extremos del conjunto de datos pero no se eliminan los valores extremos locales (inliers). Los inliers son datos que difieren significativamente de su vecindario, pero se sitúan dentro del rango general de variación del conjunto de datos.

Todos los valores atípicos (outliers e inliers) deben ser investigados y considerados como valores potencialmente erróneos antes de que se les permita permanecer como parte del conjunto de datos. Si hay pocos valores atípicos en relación con los datos completos, su eliminación frecuentemente reduce la asimetría. Luego los valores eliminados pueden ser utilizados también para la interpolación kriging. La transformación a menudo no mejora la distribución cuando hay valores extremos e incluso puede ser peor que utilizar datos sin depurar. Una alternativa recomendada es utilizar un estimador robusto como de Cressie y Hawkins (1980). Este estimador, calcula la cuarta raíz de las diferencias cuadradas y amortigua el efecto de los valores atípicos del proceso espacial.

En este trabajo se comparan dos alternativas para caracterizar la estructura de correlación espacial del rendimiento en lotes agrícolas. La primera utiliza el estimador de Matheron sobre datos depurados (eliminando outliers e inliers) y la segunda utiliza los datos sin depurar y calcula el semivariograma usando el estimador robusto de Cressie y Hawkins.

PÚBLICO DE INTERÉS:

El presente trabajo involucra a:

- Productores.
- Técnicos.
- Estudiantes.
- Comerciantes de insumos.
- Comerciantes de tecnología.
- Contratistas.
- Gobierno.
- INTA.
- Entidades privadas encargadas de elaborar estadísticas.
- Sociedad en general (personas que desean utilizar la información).

MATERIALES Y MÉTODOS

DATOS

Se utilizaron bases de datos correspondientes a 20 mapas de rendimientos con una media de 10180 registros por mapa. Los datos de rendimientos fueron registrados y georreferenciados utilizando monitores de rendimiento de lotes comerciales, ubicados al sudeste pampeano de la provincia de Buenos Aires con cultivos extensivos de mayor importancia (soja, maíz y trigo). El clima de esta región es subhúmedo-húmedo, según índice hídrico de Thornthwaite (Burgos y Vidal, 1951), con una precipitación de 880 mm por año y una temperatura media anual de 13.3°C. Los suelos predominantes pertenecen al orden de los Molisoles, gran grupo Argiudoles o Paleudoles, desarrollados sobre sedimentos loésicos, bajo régimen údico-térmico.

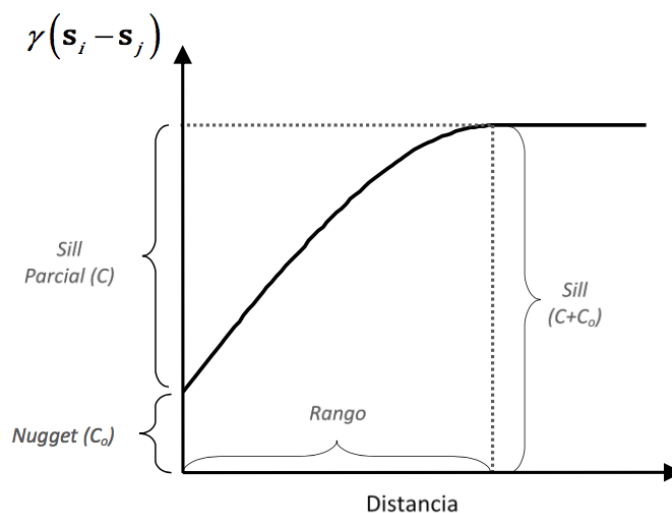
MÉTODOS

El análisis estadístico de los datos aportados por los mapas de rendimiento se realizó a través del módulo de Estadística Espacial (Córdoba y Balzarini, 2016) del software estadístico InfoStat (Di Rienzo et al., 2016). Primero para determinar el cumplimiento del supuesto de normalidad, se realizaron gráficos de distribución de frecuencias y se calcularon medidas resumen (media, mediana y coeficiente de asimetría). Posteriormente, se realizó una

depuración de datos. Para detectar y eliminar estos valores extremos (outliers e inliers) utilizamos herramientas estadísticas descriptivas (como gráficos Histogramas y Diagramas de caja (Box-plot)) y el índice de autocorrelación espacial local de Moran (IMI) (Anselini, 1995).

Para caracterizar la estructura de correlación espacial del rendimiento se calculó el semivariograma empírico (a partir de los datos). Se utilizó el estimador de Matheron sobre los datos depurados y el estimador robusto de Cressie sobre los datos sin depurar. Los parámetros de la función semivariograma usados en la comparación de estimadores fueron la varianza *nugget* o efecto pepita (C_0), la varianza estructural (C) o “sill parcial” y el rango (R). C_0 es la ordenada al origen del semivariograma, este parámetro representa la suma de errores aleatorios o no espaciales o de errores asociados con la variabilidad espacial a escalas más finas que la usada para realizar las mediciones. La varianza umbral o “sill” se obtiene sumando las varianzas antes mencionadas ($C_0 + C$) y es la varianza de observaciones independientes. El rango es el lag o distancia h en el cual la asíntota es alcanzada (Figura 1).

Figura 1. Semivariograma. Se representan los 3 parámetros que lo definen: Rango (R), Sill Parcial (C) y Nugget (C_0 .)



Realizado el ajuste del semivariograma experimental se procedió al ajuste del semivariograma teórico utilizando el método de mínimos cuadrados ponderados (WLS). Los modelos ajustados fueron el exponencial y esférico. El criterio utilizado para la elección del modelo que mejor ajustó al semivariograma empírico fue el de la suma de cuadrado del error (SCE), seleccionando el modelo que arrojó el menor valor de SCE. Posteriormente, se extrajeron los parámetros estimados con el modelo de mejor ajuste de cada mapa y con cada

estrategia de estimación. Para describir el grado de estructuración espacial se calculó la varianza estructural relativa (RSV) como:

$$RSV = \left(\frac{C}{C + C_0} \right) \times 100\%$$

Para comparar los estadísticos resumen (Media y CV) y parámetros estimados del semivariograma (Sill Parcial, Nugget, Rango y RSV) con cada método de estimación, se utilizó la prueba no paramétrica de Wilcoxon para muestras apareadas (Mann-Whitney U).

RESULTADOS Y DISCUSIÓN

En el análisis descriptivo de los datos, tanto de la base original (sin depurar) como de los datos depurados, se observó una diferencia entre las medias, en el que los datos no depurados tuvieron mayor rendimiento promedios a través de los mapas analizados. La mayor diferencia se visualizó a nivel del CV, siendo la misma significativa y mayor para los datos sin depurar (133% vs. 48%) (Tabla 1).

Tabla 1 Medias resumen. Media, Mínimo y Máximo del rendimiento promedio y CV de 20 mapas de rendimiento y diferencias entre Medias de rendimiento y Coeficiente de Variación (CV%) obtenidos mediante la prueba Wilcoxon. Datos depurados y no depurados con sus respectivos p-valor.

	Rendimiento promedio (Tn/ha)		Coeficiente de variación (%)	
	No depurados	Si depurados	No depurados	Si depurados
Media	4,1	3,8	132,9	47,8
Mínimo	1,1	1,5	35,2	31,9
Máximo	10,8	10,1	574,9	69,7
Diferencia promedio (depurados - sin depurar)		-0,31		-85
p-valor		0,001		0,0006

En el proceso de ajuste de semivariogramas, utilizando la base de los 20 mapas de rendimientos depurados se obtuvieron los respectivos 20 semivariogramas con sus parámetros, no así sobre los mapas de rendimiento no depurados, donde en un sólo mapa no se pudo realizar la estimación del semivariograma. Los parámetros del semivariograma (nugget, sill parcial, rango) y el RSV se compararon mediante la prueba de Wilcoxon. Los resultados mostraron que no se encontraron diferencias significativas para el efecto *nugget* (C_0) y sill parcial (C) ($p > 0.05$), siendo sus valores menores en los datos sin depurar bajo el estimador de Matheron. Dado que el (n) para la comparación de medias es bajo ($n=20$), probablemente utilizando una mayor cantidad de mapas, las diferencias en estos parámetros hubiesen sido significativas aún con un nivel de significación del 0.05 (Tabla 2). Para el parámetro rango (R) sí se observó una diferencia significativa, ($p=0,0466$) siendo la misma en promedio de 559,6 m mayo en los datos sin depurar usando el estimador de Matheron.

Un mayor rango implicaría mayor contigüidad espacial, buscado en el contexto de la agricultura de precisión para la definición de zonas homogéneas. Sin embargo, cuando no se depuran los datos y se usa el estimador robusto de Cressie los rangos estimados pueden tomar valores muy por encima del esperado para el dominio estudiado, es decir, se estiman rangos a una distancia mayor a la máxima que puede existir entre los puntos muestreados. Esto puede hacer que la variabilidad espacial intralote no se observe.

Tabla 2. Diferencias promedios de parámetros de semivariogramas teóricos ajustados y grado de estructuración espacial (RSV) con sus respectivos p-valor obtenidos mediante la prueba de Wilcoxon. Datos depurados con estimador de Matheron (Si Depurados) y datos no depurados con estimador de Cressie (No Depurados).

Parámetros estimados del semivariograma	Si Depurados – No Depurados	p-valor
Nugget	-29,4	0,0662
Sill parcial	-70,68	0,0582
Rango	-559,67	0,0466
RSV (%)	-7	0,0244

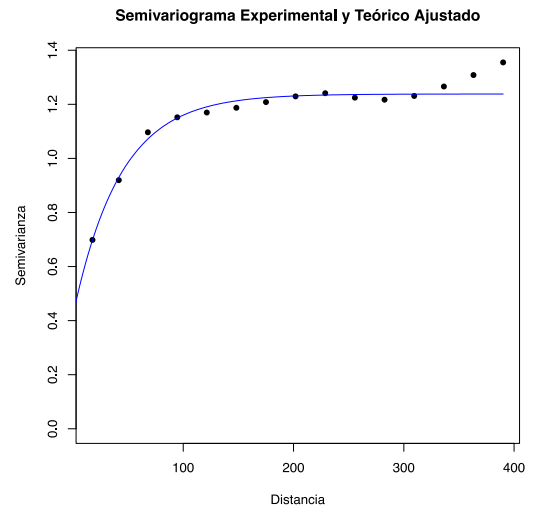
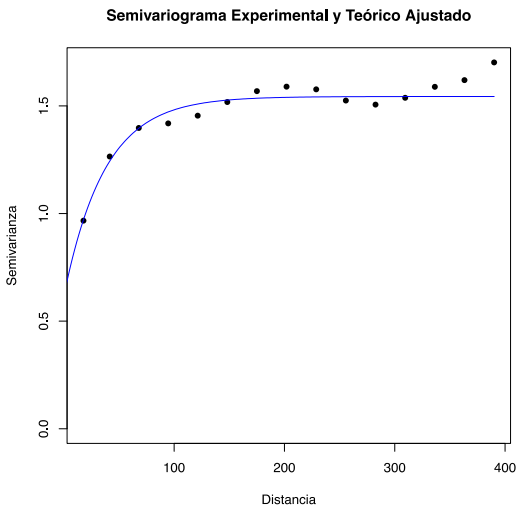
Cuando se analiza el RSV la diferencia es levemente menor del 7%, es decir el grado de estructuración espacial utilizando los datos no depurados con el estimador de Cressie, es

mayor respecto a los obtenidos con el estimador de Matheron a partir de datos depurados, siendo esta diferencia significativa ($p = 0,0244$). La poca diferencia en la RSV se debe a que las diferencias importantes en el semivariograma se dan en todos los parámetros, pero no así en la relación entre ellos, esta se mantiene similar.

Respecto a los modelos de semivariograma teóricos seleccionados según la SCE, usando el estimador de Martheron y datos depurados 16 mapas se ajustaron al modelo exponencial mientras que los 4 restantes el modelo de mejor ajuste fue el esférico. Ahora bien, cuando la estimación del semivariograma se realizó sobre la base de datos no depurado con el estimador de Cressie, el número de modelos no tuvo un cambio importante, de los 19 semivariogramas ajustados (para un mapa no fue posible realizar la estimación), 16 se correspondieron al modelo exponencial y los 3 restantes al esférico. En sólo un mapa se produjo un cambio del modelo ajustado. Cuando se usaron datos no depurados y el estimador de Cressie el modelo de mejor ajuste fue exponencial y luego de la depuración cambió al esférico.

Utilizando las bases de datos en las cuales se encontraron las mayores y menores diferencias entre parámetros del semivariograma se realizó la predicción espacial y el mapeo de la variabilidad espacial (rendimiento), utilizando una grilla de 10 x 10. En los semivariogramas con mayores diferencias en los parámetros estimados, se observan cambios importantes a nivel de la varianza sill y en el rango estimados. Cuando los datos no se depuran y se utiliza el estimador de cressie, el sill y el rango aumentaron (Figura 2). En los mapas de rendimiento, se observa que no existen diferencias importantes en los patrones de los mapas de variabilidad espacial. Esto podría indicar que, en el contexto de un proceso de delimitación de zonas multivariado, tal vez no sea tan importante la depuración y selección de un método de estimación de semivariograma robusto, siempre y cuando los patrones de variación se mantengan y no se vean alterados. Sera necesario más estudios para evaluar su impacto en dicho proceso.

a)



b)

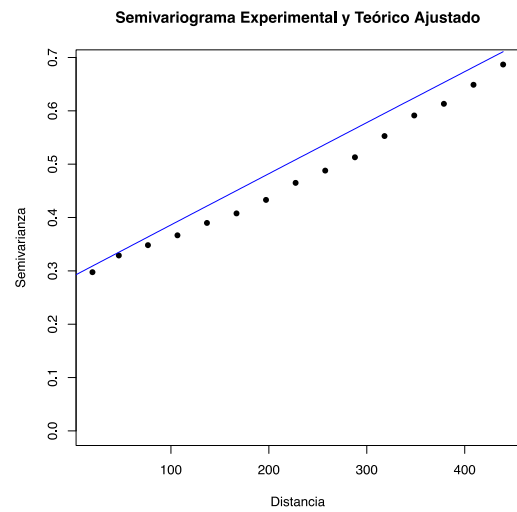
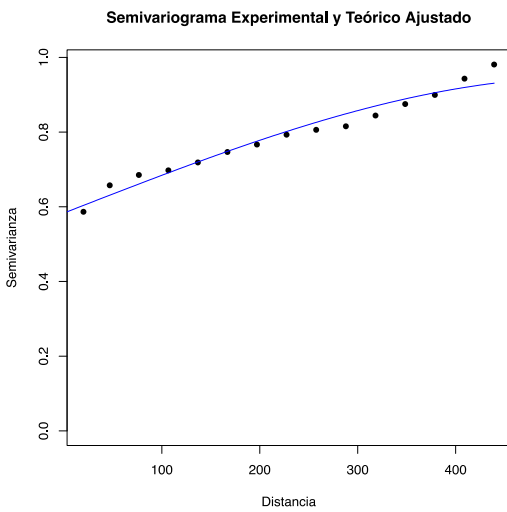
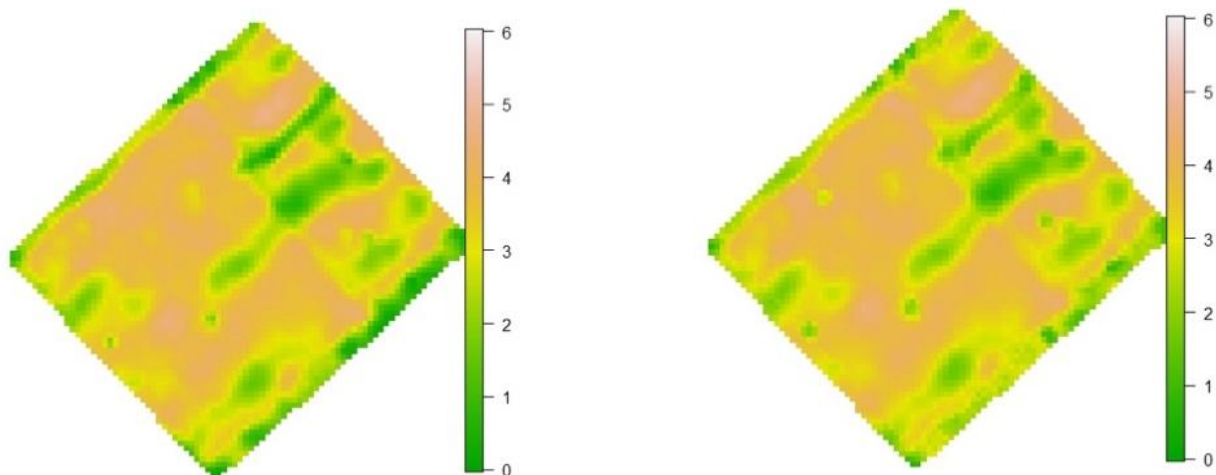


Figura 3: Semivariogramas empíricos (círculos) y teóricos (línea) obtenidos a partir del estimador robusto de Cressie para datos sin depurar (izquierda) y el estimador de Matheron para datos depurados (derecha). a) Ejemplo de semivariograma con menores diferencias entre parámetros estimados. b) Ejemplo de semivariograma con mayores diferencias entre parámetros estimados.

a)



b)

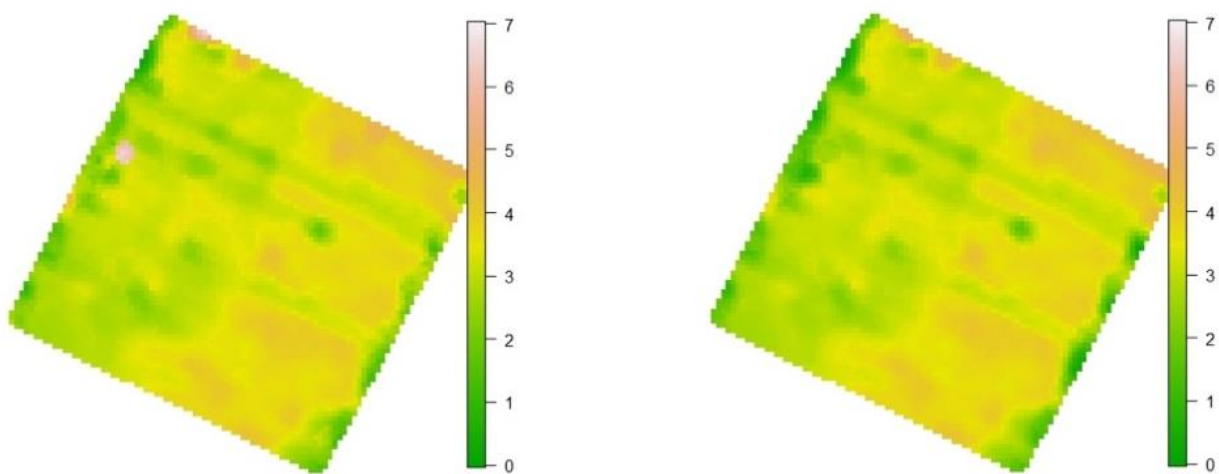


Figura 4. Mapas de variabilidad espacial de rendimiento obtenidos a partir de parámetros de semivariogramas estimados a partir del estimador robusto de Cressie para datos sin depurar (izquierda) y el estimador de Matheron para datos depurados (derecha). a) Ejemplo de semivariograma con menores diferencias entre parámetros estimados. b) Ejemplo de semivariograma con mayores diferencias entre parámetros estimados.

CONCLUSIÓN

Entre datos depurados ajustados con el estimador clásico y datos no depurados ajustados con el estimador robusto, si bien se observa una importante diferencia en los parámetros que caracterizan la estructura de correlación espacial, la relación entre estos parámetros y los patrones de variación se mantuvieron similares.

CONSIDERACIONES FINALES

A pesar de los beneficios obtenidos a partir de la implementación de tecnología¹² de agricultura de precisión como lo es el monitor de rendimiento, el correcto análisis y manejo de los datos es una etapa crucial para la implementación de estas tecnologías, evitando fallas¹³ en la toma de decisiones. El gran volumen de datos aportados por los monitores de rendimiento requiere de métodos de depuración o limpieza para que el análisis de la variabilidad espacial dentro del lote refleje los patrones reales¹⁴ del lote.

Por último podemos agregar que si bien el trabajo planteado fue de investigación¹⁵, la implementación de estas metodologías sería de suma importancia en situaciones de la realidad, pudiendo añadir estos datos juntos con otros de importancia agronómica (fertilización, densidades, contenido de materia orgánica¹⁶) potenciando de esta forma esta nueva herramienta tecnológica.

¹² Indicador 11: Competencia leal.

¹³ Indicador 31: Relacionamiento con el consumidor.

¹⁴ Indicador 12: Prácticas anti corrupción.

¹⁵ Publico de interés.

¹⁶ Indicador 41: Uso sustentable de los recursos materiales.

BIBLIOGRAFÍA

- Anselini, Luc. Local Indicators of Spatial Association-LISA. Geographical Analysis. Ohio: Ohio State University Press, 1995, vol. 27, n° 2, pp. 93-115.
- Balzarini M.G., Gonzalez L., Tablada M., Casanoves F., Di Rienzo J.A., Robledo C.W. 2008. Infostat. Manual del Usuario, Editorial Brujas, Córdoba, Argentina.
- Bongiovanni R., Montovani E.C., Best S. y Roel, A. 2006. Agricultura de precisión: Integrando conocimientos para una agricultura moderna y sustentable. PROCISUR/ IICA, Montevideo, 246 pp.
- Burgos, J.J. y A.L. Vidal, 1951. Los climas de la República Argentina según la nueva clasificación de Thornthwaite. Meteoros 1:3-33.
- Bullock D.S. and Lowenberg-DeBoer J. 2007. Using Spatial Analysis to Study the Values of Variable Rate Technology and Information. JAE 58(3): 517–535.
- Cressie N. A. C. 1993. Statistics for spatial data. New York: J. Wiley and Sons, Estados Unidos 900 pp.
- Cressie N. A. C.; Hawkins, D. M. Robust estimation of the variogram. Mathematical Geology, New York, v.12, n.2, p.115-125, 1980.
- Córdoba M., Balzarini M. 2016. Análisis de datos espaciales en el software InfoStat. III Jornadas Académicas de Estadística – Geoestadística - Sociedad Argentina de Estadística. Neuquén. Agosto de 2016.
- Di Rienzo J.A., Casanoves F., Balzarini M.G., Gonzalez L., Tablada M., Robledo C.W. InfoStat versión 2016. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina.
- Indicadores Ethos. 2008. Indicadores Ethos de Responsabilidad Social Empresarial. 78 pp.
- Martinez Melo, C.E. 2010. Análisis Geoestadístico Espacio Tiempo Basado en Distancias y Splines con Aplicaciones, Universidad Autónoma de Barcelona, España.
- Matheron G. The theory of regionalized variables and its applications. Cahiers du Centre de Morphologie Mathematique, 5. Fontainebleau, France. 1971.
- Schabenberger O. and Gotway C. A. 2004. Statistical Methods for Spatial Data Analysis. Taylor and Francis. Chapman and Hall/CRC, 488pp.
- Webster R. and Oliver M.A. 2007. Geostatistics for environmental scientists, 2nd edn. John Wiley and Sons, Chichester UK.