

Efecto de la definición de vecindario en la depuración de mapas de rendimiento

CISNEROS MARIO

GODOY RODRIGO

QUINTERO DIEGO

PAÑART SANTIAGO

TUTOR: Ing. Agr. Dr. Mariano Córdoba

AÑO 2016

Introducción

Desde los principios de la agricultura la variabilidad espacial del suelo ha estado presente en la producción agrícola. Una de las tendencias globales en la agricultura moderna es buscar la eficiencia del sistema actuando sobre el espacio productivo de manera diferenciada, es decir, en lugar de realizar manejos uniformes o promedios del lote se busca el manejo por unidades ambientales que justifiquen su diferenciación en cuanto a que son susceptibles de diferentes manejos ó prácticas agronómicas¹. En este contexto ha tenido una amplia difusión el concepto de agricultura de precisión (AP). La definición más simple, establece que corresponde a un grupo de tecnologías que permiten la aplicación de insumos agrícolas, tales como fertilizantes, semillas, plaguicidas, entre otros, en forma variable dentro de un lote de acuerdo a los requerimientos y/o potencial productivo de zonas homogéneas intralote previamente delimitadas (Ortega y Flores, 1999).

Existen numerosas tecnologías de precisión que se consideran forman parte de la AP en sus diferentes etapas de aplicación. Entre ellas se destacan herramientas como el GPS, sensores remotos, sensores proximales de suelo y del cultivo, que se utilizan para la recolección de información en múltiples sitios dentro del lote. Una tecnología de amplio uso y que se considera una de las principales fuentes de información son los mapas de rendimiento. Los monitores de rendimiento acoplados a los equipos de cosecha, permiten recabar datos georreferenciados de rendimiento produciendo gran cantidad de datos en áreas relativamente pequeñas y en cortos períodos de tiempo (Córdoba *et al.*, 2015). Estos datos son posteriormente utilizados para la generación de mapas digitales de rendimiento que permiten visualizar la variabilidad espacial del rendimiento intralote o bien marcar en forma simple los factores de manejo y su respuesta en el rendimiento. Esta capa de información frecuentemente es utilizada como input para la delimitación de zonas de manejo (ZM) (Córdoba *et al.*, 2016). Las ZM son áreas dentro del lote que se caracterizan por presentar características similares, tales como textura, topografía, estado hídrico y niveles de nutrientes (Moral *et al.*, 2010).

La recolección de datos de rendimiento durante la cosecha tiene como ventaja el bajo costo y la gran cantidad de puntos de muestreo (aproximadamente 250 por hectárea, dependiendo del ancho de corte y la frecuencia de grabado). Sin embargo, debido a que es un proceso automático que se realiza en forma secundaria a la operación de cosecha, es susceptible a diversas fuentes de

¹ TECNOSEM.com.ar/ Los mapas de rendimiento como una fuente de información útil para la gestión agronómica

error. Algunos de esos errores se originan por fallas en la preparación y mantenimiento de los componentes del sistema que toma los datos (GPS, consola, sensores) o bien, son inducidos por el operario durante el proceso de cosecha². Ejemplos de estos últimos son: la indebida calibración de peso y humedad, incorrecto ajuste del ancho de corte. Otros, más que errores son desviaciones en la medición, y ocurren producto de la manera en que funciona el sistema de mapeo en sí, como por ejemplo, atoro de la cosechadora, entradas y salidas de la cabecera, cosecha con más de una máquina y errores de GPS. Es por ello que es recomendable realizar un pre-procesamiento de los datos antes de disponerse a analizarlos ó bien a utilizarlos como base para determinar conclusiones productivas o decisiones futuras de manejo. Dicho pre-procesamiento incluye una depuración para poder eliminar datos raros o extremos (outliers e inliers).

Los outliers son observaciones con valores extremos 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 (Taylor *et al.*, 2007). Una forma de eliminar los outliers es a partir de datos de la variable rendimiento, calcular la media y la desviación estándar (DE) para identificar los valores que se encuentran fuera de la media ± 3 DE (Córdoba *et al.*, 2014). Según la teoría, el 99% de los datos se van a encontrar entre la media ± 3 SD cualquiera sea la distribución de la variable. Por otro lado, los inliers, o valores extremos locales (outliers espaciales), son datos que difieren significativamente de su vecindario pero se sitúan dentro del rango general de variación del conjunto de datos (Córdoba *et al.*, 2014). Para la identificación y eliminación de dichos valores se utilizó el índice autocorrelación espacial local de Moran (IMI), el cual es aplicado a cada dato individualmente y da idea del grado de similitud o diferencia entre el valor de una observación respecto al valor de sus vecinos. (Anselin, 1995; Córdoba *et al.*, 2016). Se considera puntos vecinos de un sitio determinado a aquellos contiguos ubicados entre un rango de distancia definido por un límite inferior y un límite superior previamente preestablecido.

Hasta el momento no se conocen estudios sobre el efecto que tiene la definición de los vecindarios para la eliminación de inliers en la caracterización de la estructura de variabilidad espacial del mapa de rendimiento. Particularmente sobre los parámetros del semivariograma (Crescié, 1993), función utilizada para la interpolación espacial y generación del mapa de variabilidad espacial (Webster y Oliver, 2007).

² TECNOSEM.com.ar/ Los mapas de rendimiento como una fuente de información útil para la gestión agronómica

En este trabajo se analiza el efecto que tiene la definición de los vecindarios, utilizando diferentes distancias máximas, sobre parámetros del semivariograma y estadísticos descriptivos de la variable rendimiento.

Materiales y Métodos

Bases de datos

Se utilizaron 35 mapas de rendimiento de lotes agrícolas ubicados en el sudoeste de la provincia de Buenos Aires. Cada base contó con los datos de posición geográfica y el valor del rendimiento para distintos cultivos de grano (soja, maíz, trigo, cebada, girasol). La obtención de estos datos se realizó mediante la utilización de un sistema de posicionamiento global (GPS), el cual determina la posición espacial mediante un sistema de coordenadas geodésicas, es decir, expresados en grados, minutos y segundos. Para facilitar los cálculos de distancia y superficie, se realizó una transformación de dichas coordenadas, a coordenadas cartesianas, la cual permite que las distancias entre sitios o puntos desde donde se leen los datos, se expresen como una distancia absoluta (metros), en vez de distancias relativas (grados), realizando así una mejor interpretación de los análisis.

Depuración de datos

En primera instancia se eliminaron datos que presentasen rendimientos inferiores a 0,1 qq/ha como umbral biológico mínimo según rendimientos históricos esperados en el área agrícola.

El siguiente paso consistió en la depuración de los datos, es decir identificación y eliminación de Outliers e Inliers. Primero se eliminaron los outliers, es decir, aquellos datos que se ubicaron por fuera del intervalo $\text{media} \pm 2,5 \text{ DE}$. Luego para eliminar los Inliers se calculó el IMI usando como límites superiores una secuencia de distancias entre 10 y 100 m con intervalos de 5 m. Así, se obtuvieron 35 (mapas) \times 19 (distancias) bases depuradas, las cuales fueron concatenadas en una misma tabla para el posterior análisis conjunto.

Estadística descriptiva y ajuste de semivariograma

Se realizó la estadística descriptiva y se obtuvo el semivariograma empírico, y sobre éste, se ajustaron por el método WLS, distintos modelos de semivariograma. Se evaluaron los modelos exponencial, esférico y gaussiano. La suma de cuadrado del error (SCE) fue el criterio usado para la selección del modelo que mejor ajustó el semivariograma empírico. Para cada mapa y distancia usada en la depuración, se calcularon la media, mínimo, máximo, mediana, CV, sill parcial, nugget, rango y el porcentaje de varianza estructural relativa (RSV). La varianza nugget o efecto pepita, es la ordenada al origen del semivariograma (Fig. 1) y 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. Un alto valor de C_0 indica que la mayoría de la variación ocurre en distancias más cortas que la mínima distancia que separa dos observaciones en la grilla de estudio. La varianza estructural o parcial sill es la asíntota, y la varianza umbral o sill, se obtiene sumando las varianzas antes mencionadas y es la varianza de observaciones independientes.

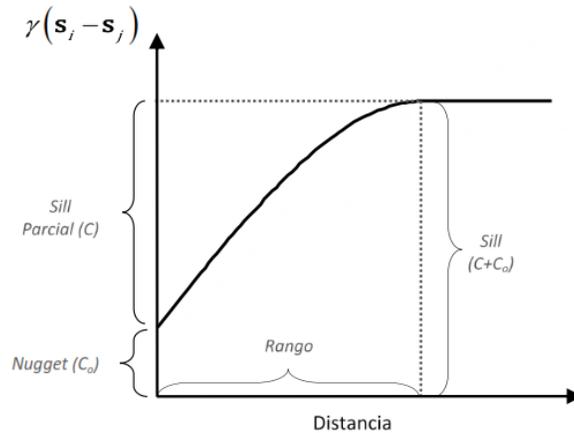


Fig. 1. Semivariograma. Se representan los tres parámetros que lo definen: rango, sill y nugget.

El RSV permite cuantificar el grado de estructuración espacial (Eq. 1). Un alto valor de RSV indica que las predicciones geoestadísticas serán más eficientes que aquellas obtenidas con métodos de predicción que ignoran la información espacial.

$$(Eq. 1) \quad RSV = \frac{C}{C+C_0} \times 100$$

C_0 : La varianza *nugget* o efecto pepita.

C : La varianza estructural o *partial sill*.

$C_0 + C$: La varianza umbral.

Los estadísticos descriptivos y parámetros del semivariograma de cada mapa y cada distancia máxima usada en la depuración, fueron registrados en una misma tabla (Fig. 2). También a los datos obtenidos fueron estandarizados por el máximo de cada mapa (SCMX_n, SCMX_sill parcial, SCMX_CV, SCMX_rango, SCMX_media, SCMX_RSV, SCMX_Min, SCMX_Max, SCMX_Nugget), para así obtener valores relativos entre 0 a 1, y comparar los resultados entre los diferentes mapas (Fig. 3).

Caso	Mapa	Distancia	Modelo	Rango	psill	Nugget	n	Media	Min.	Max.	CV
1	Mapa 1	10,00	"Exp"	25,31	0,20	0,08	3771,00	2,97	0,77	4,81	19,34
2	Mapa 1	15,00	"Exp"	31,59	0,19	0,09	3766,00	2,97	0,77	4,81	19,24
3	Mapa 1	20,00	"Exp"	38,79	0,19	0,10	3756,00	2,97	0,77	4,81	19,18
4	Mapa 1	25,00	"Exp"	49,42	0,19	0,10	3739,00	2,97	0,77	4,81	19,04
5	Mapa 1	30,00	"Exp"	57,87	0,19	0,10	3725,00	2,97	0,77	4,81	18,96
6	Mapa 1	35,00	"Exp"	65,27	0,19	0,10	3714,00	2,97	0,77	4,81	18,89
7	Mapa 1	40,00	"Exp"	72,44	0,19	0,11	3703,00	2,97	0,77	4,81	18,90
8	Mapa 1	45,00	"Exp"	74,98	0,19	0,11	3691,00	2,97	0,77	4,81	18,87
9	Mapa 1	50,00	"Exp"	79,89	0,19	0,10	3677,00	2,97	0,77	4,81	18,80
10	Mapa 1	55,00	"Exp"	80,03	0,19	0,10	3675,00	2,97	0,77	4,81	18,79
11	Mapa 1	60,00	"Exp"	78,54	0,19	0,10	3667,00	2,97	0,77	4,81	18,73
12	Mapa 1	65,00	"Exp"	79,76	0,20	0,10	3654,00	2,97	0,77	4,81	18,73
13	Mapa 1	70,00	"Exp"	79,97	0,20	0,09	3641,00	2,97	0,77	4,81	18,62
14	Mapa 1	75,00	"Exp"	77,68	0,20	0,09	3635,00	2,97	0,77	4,81	18,62
15	Mapa 1	80,00	"Exp"	80,94	0,20	0,09	3628,00	2,97	0,77	4,81	18,58
16	Mapa 1	85,00	"Exp"	80,47	0,20	0,09	3624,00	2,97	0,77	4,81	18,56
17	Mapa 1	90,00	"Exp"	78,29	0,20	0,08	3618,00	2,97	0,77	4,81	18,57
18	Mapa 1	95,00	"Exp"	80,70	0,20	0,08	3606,00	2,97	0,77	4,81	18,60
19	Mapa 1	100,00	"Exp"	83,44	0,20	0,09	3603,00	2,97	0,77	4,81	18,62

Real Registros: 19*11 n=1 Suma = 19,34 Media = 19,343 D.E. = 0,00 Min = 19,34 Max = 19,34 P05 = 19,34 P95 = 19,34

Fig. 2. Estadísticos descriptivos y parámetros del semivariograma para el mapa 1 y distancias máximas usadas para la definición de vecindario de cada dato entre 10 a 100 m.

Caso	Max.	CV	RSV	SCMx_Rango	SCMx_psil	SCMx_Nugget	SCMx_n	SCMx_Media	SCMx_Min.	SCMx_Max.	SCMx_CV	SCMx_RSV
651	5,99	13,54	54,55	0,55	1,00	0,88	0,99	1,00	1,00	1,00	0,98	1,00
652	5,99	13,53	53,70	0,58	0,98	0,89	0,99	1,00	1,00	1,00	0,98	0,98
653	5,99	13,54	52,81	0,62	0,97	0,91	0,99	1,00	1,00	1,00	0,98	0,97
654	5,99	13,51	52,64	0,64	0,96	0,91	0,98	1,00	1,00	1,00	0,98	0,96
655	5,99	13,53	51,42	0,68	0,95	0,94	0,98	1,00	1,00	1,00	0,98	0,94
656	5,99	13,47	51,59	0,70	0,94	0,93	0,98	1,00	1,00	1,00	0,98	0,95
657	5,99	13,45	51,40	0,76	0,93	0,93	0,97	1,00	1,00	1,00	0,98	0,94
658	5,99	13,44	50,87	0,78	0,92	0,94	0,97	1,00	1,00	1,00	0,98	0,93
659	5,99	13,41	51,38	0,79	0,92	0,92	0,97	1,00	1,00	1,00	0,97	0,94
660	5,99	13,42	50,81	0,83	0,92	0,93	0,96	1,00	1,00	1,00	0,97	0,93
661	5,99	13,43	49,82	0,88	0,90	0,95	0,96	1,00	1,00	1,00	0,98	0,91
662	5,99	13,45	49,17	0,92	0,89	0,97	0,96	1,00	1,00	1,00	0,98	0,90
663	5,99	13,45	48,93	0,93	0,88	0,97	0,95	1,00	1,00	1,00	0,98	0,90
664	5,99	13,49	47,62	1,00	0,86	1,00	0,95	1,00	1,00	1,00	0,98	0,87
665	5,99	13,45	48,16	0,99	0,87	0,98	0,95	1,00	1,00	1,00	0,98	0,88

Real Registros: 665*22 n=1 Suma = 0,88 Media = 0,883 D.E. = 0,00 Min = 0,88 Max = 0,88 P05 = 0,88 P95 = 0,88

Fig. 3. Estadísticos descriptivos y parámetros del semivariograma estandarizados por el máximo.

Todos los análisis fueron realizados utilizando el módulo Estadística Espacial (Córdoba y Balzarini, 2016) del software InfoStat (Di Rienzo et al., 2016).

Resultados

La media, mínimo, máximo y percentil 50 de los CV promedios de los parámetros del semivariograma (Rango, Psill, Nugget) y estadísticos descriptivos (n, Media, Min, Max, RSV), se muestran en la Tabla 1.

Tabla 1. Estadísticos descriptivos y parámetros del semivariograma.

Resumen	Rango	psill	Nugget	n	Media	Min.	Max.	RSV ³
Media	31,4	7	10	2,7	0,3	2,9	0,2	4,1
Mín	11,5	1,2	2,6	1,2	4,00E-02	0	0	0,7
Máx	63,4	38,5	41,8	4,6	1,5	71	2,4	11,4
P(50)	28,1	5,5	7,8	2,6	0,2	0	0	3,8

Respecto a los parámetros del semivariograma, el Rango fue el que más variabilidad tuvo (CV=32), seguido por el efecto nugget. Se observó una relación significativa y positiva entre la distancia y el

$$^3 RSV = \frac{psill}{(psill+nugget)}$$

rango estandarizado (Fig. 4). Los restantes parámetros del semivariograma y el RSV no presentaron una tendencia sistemática en todos los mapas. Por ejemplo, para el efecto nugget algunos mapas tuvieron una relación positiva, negativa o no cambiaron con la distancia (Fig. 5).

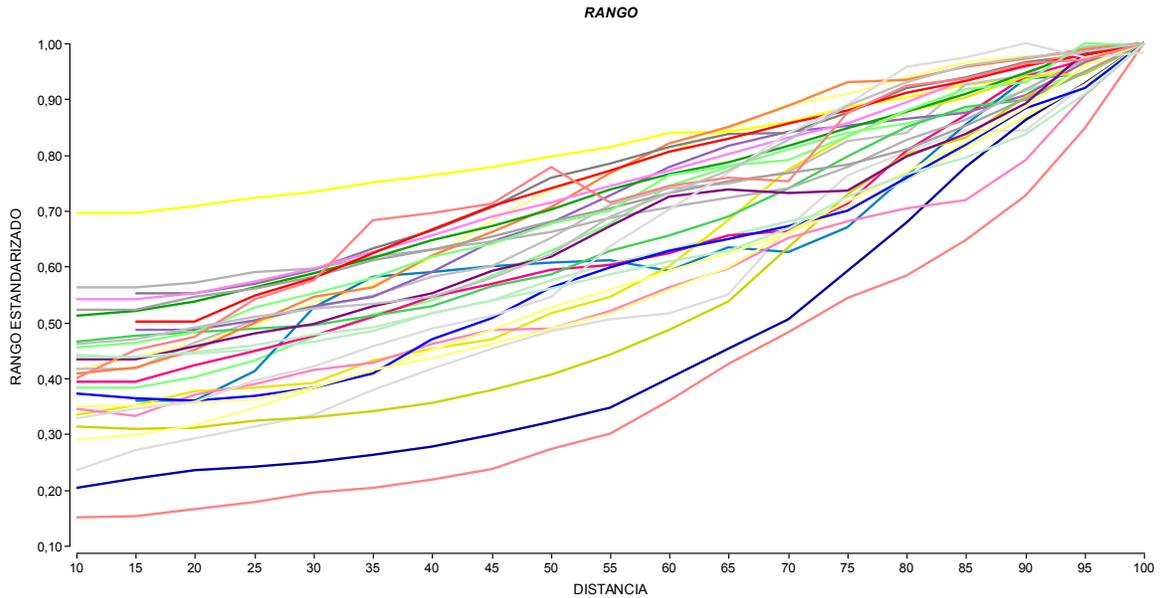


Fig. 4. Relación entre el rango estandarizado por el máximo y la distancia máxima usada para la definición del vecindario.

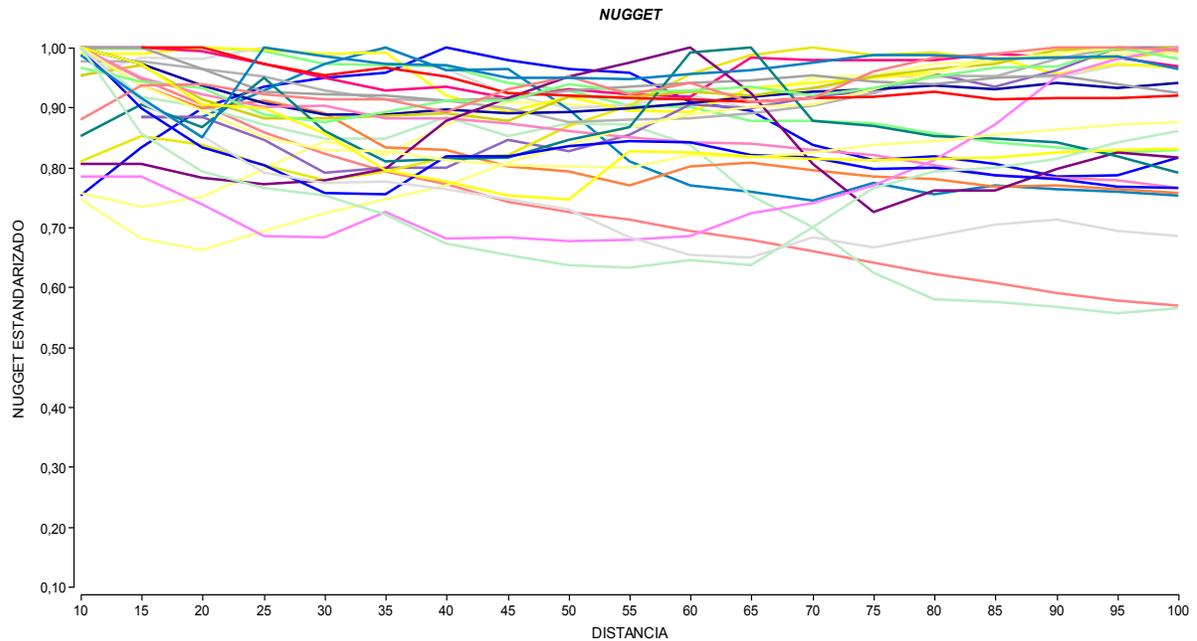


Fig. 5. Relación entre el nugget estandarizado por el máximo y la distancia máxima usada para la definición del vecindario.

En cuanto a los estadísticos descriptivos, el **n** tuvo uno de los mayores valores de CV aunque comparado con los del semivariograma esta fue baja (CV=3). Para este estadístico también se observó una relación significativa con la distancia máxima usada en la definición de los vecindarios de cada dato. El **n** mismo disminuyó a medida que aumentó la distancia, esto significa que se eliminaron más datos a medida que aumento la distancia (Fig. 6). En cuanto al Max. Y Min. sus valores no se vieron modificados al aumentar la distancia.

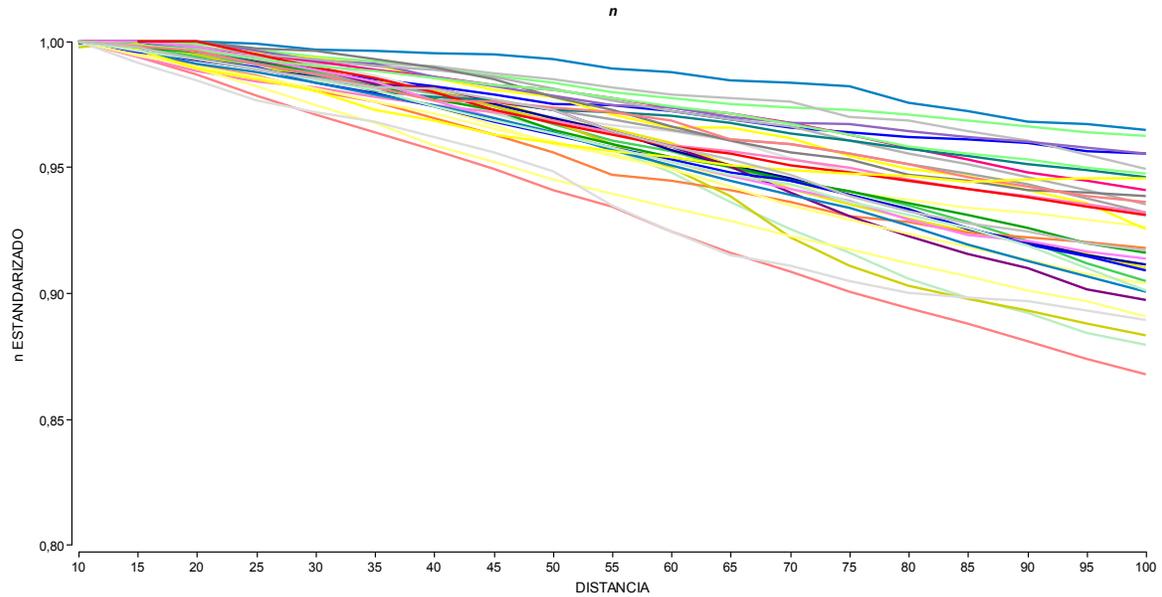
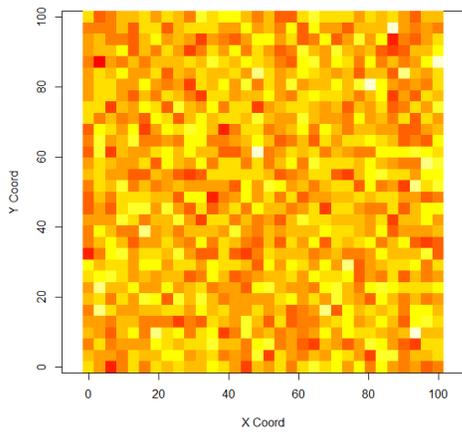


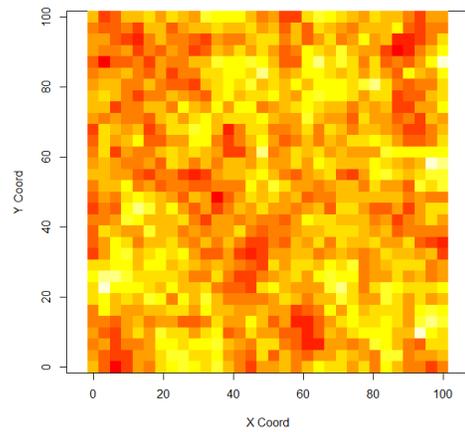
Fig. 6. Relación entre el n estandarizado por el máximo y la distancia máxima usada para la definición del vecindario.

Conclusiones

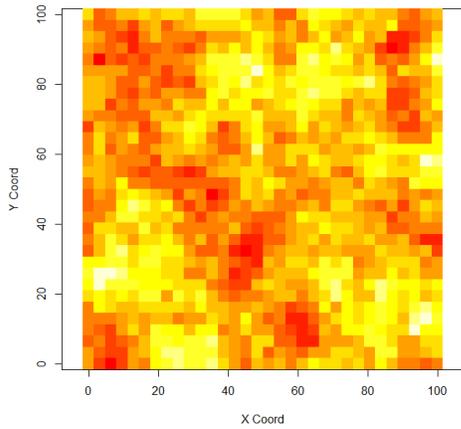
Los resultados mostraron que tanto los parámetros del semivariograma como los estadísticos descriptivos utilizados, varían de distinta manera a medida que aumenta la distancia máxima usada en la definición de vecindarios, paso necesario para el cálculo del IMI. El rango fue el parámetro del semivariograma que tuvo mayor variabilidad y presentó una relación positiva con la distancia. Mientras que para los descriptivos el n presentó una relación negativa con la distancia, aunque su variabilidad fue baja. Esto implica que a medida que se aumenta la distancia se logra una mayor contigüidad espacial debido a que se elimina una mayor cantidad de datos, siendo los cambios en el n de escasa magnitud. Con ello se logran interpolaciones más suavizadas que podrán facilitar la implementación de manejos diferenciados (Fig. 7).



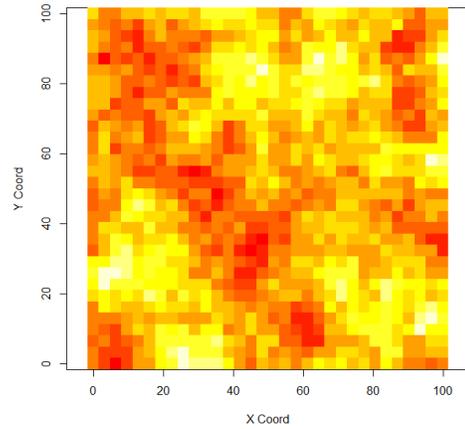
Rango: 1



Rango: 5



Rango: 10



Rango: 15

Fig. 7. Cambio en la contigüidad de un mismo mapa de rendimiento a medida que aumenta el rango.

Referencias bibliográficas

Amuchastegui J. 1999. "interpretación de mapas de rendimientos". Publicado en internet, disponible en www.inta.org.ar. Activo Octubre 2016.

Anselin L., (1995). Local indicators of spatial association – LISA. *Geographical Analysis*, 27: 93-115.

Córdoba M., Bruno C., Aguate F., Tablada M., Balzarini M. (2014). Análisis de la variabilidad espacial en lotes agrícolas. Manual de Buenas Prácticas. Ed. Balzarini, M. Eudecor. Córdoba, Argentina. pp.8 – 9.

Córdoba M. (2015). Estadística y Biometría. Ed. Brujas, Córdoba, Argentina. pp. 335 - 350. ISBN 978987-591-646-3.

Córdoba M., Balzarini M. 2016. Análisis de datos espaciales en el software InfoStat. III Jornadas Académicas de Estadística – Geoestadística. Sociedad Argentina de Estadística. Neuquen, Agosto de 2016.

Córdoba M, Bruno C., Costa J., Peralta N., Balzarini M. (2016). Protocol for multivariate homogeneous zone delineation in precision agriculture. Biosyst. Eng. 143: 95-107.

Córdoba M., Bruno C., Costa J. L., Balzarini M. (2016). Variabilidad espacial multivariada en variables de suelo a escala de lote y su relación con rendimientos. Revista de Investigaciones Agropecuarias. 42: 47:53.

Cressie N. A. C., (1993). Statistics for Spatial Data Revised Edition. John Wiley and Sons, New York, 900 pp.

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. URL <http://www.infostat.com.ar>

Mendez A. (2010). Puntos clave para lograr un mapa de rendimiento con datos confiables. Publicado en internet, disponible en www.inta.org.ar. Activo Octubre 2016.

Moral F. J., Terrón J.M. and 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.

Oliver M. A., (2013). Precision agriculture and geostatistics: How to manage agriculture more exactly. *Significance*, 10(2): 17–22.

Ortega, R. y Flores L. (1999). Agricultura de precisión. p13-46. En R. Ortega y L. Flores (ed.) *Agricultura de Precisión: Introducción al manejo sitio - específico*. Ministerio de Agricultura, Instituto de Investigaciones Agropecuarias, CRI Quilamapu-Chile.

Taylor J. A., McBratney A. B. and Whelan B. M., (2007). Establishing Management Classes for Broadacre Agricultural Production. *Agron. J.* 99: 1366–1376.

Webster R. and Oliver M. A., (2007). *Geostatistics for environmental scientists*, 2nd edn. John Wiley and Sons, Chichester UK.