

La Econometría Espacial en la Agricultura de Precisión: Aplicación en un ingenio azucarero

¹Vítores Sánchez, X. A.; ²Sandoya Sánchez, F.

¹Ingeniero en Estadística e Informática, xavi2426@hotmail.com

²Matemático, M.Sc. Operaciones y Logística, fsandoya@espol.edu.ec

Escuela Superior Politécnica del Litoral ESPOL
Instituto de Ciencias Matemáticas ICM

Campus Gustavo Galindo, Km 30½ Vía Perimetral, Prosperina, Guayaquil, Ecuador

RESUMEN

La Agricultura Precisión (AP) es el uso de tecnologías de información geográfica para adecuar el manejo de suelos y cultivos a la variabilidad presente dentro de la unidad de manejo agrícola. Esta filosofía de un manejo sitio-específico del cultivo nace como respuesta al tradicional manejo uniforme de las parcelas que no aprovecha todo el potencial de los recursos naturales y económicos utilizados. Existe una brecha de información entre análisis de datos espaciales de las características del suelo y cultivo y las recomendaciones sitio-específicas que no ha permitido una rápida adopción de la AP. La correlación espacial entre observaciones vecinas no se adapta a los supuestos del análisis estadístico clásico, lo que se traduce en conclusiones incorrectas. Esta investigación explora la utilización de la econometría espacial para estimar una función de la respuesta sitio-específica de un cultivo de caña de azúcar con respecto a sus principales factores determinantes. Los beneficios de la econometría espacial se derivan de la incorporación explícita de los efectos espaciales de los datos georeferenciados dentro de un modelo matemático. Se estableció un diagnóstico de la situación del cultivo, se pudieron realizar inferencias válidas y se pudieron establecer recomendaciones de manejo sobre una base estadística robusta.

Palabras clave: Manejo sitio-específico, agricultura de precisión, econometría espacial, tecnologías de información geográfica.

ABSTRACT

Precision Agriculture (PA) is the application of geographical information technologies in order to be able to satisfy crop and soil requirements related to the crop management unit variability. This philosophy of a site-specific crop management appears as an alternative to traditional uniform crop management, which doesn't take advantage of the whole natural and economic resources potential. One of the key constraints on adoption of PA is the gap between data analysis and site-specific recommendations. Analysis of spatial data has proven difficult because correlation among neighboring observations often violates the assumptions of classical statistical analysis. This research surveys the use of spatial econometrics to estimate a sugar cane crop yield response function related to its principal production factors. The benefits of spatial econometrics are in that spatial effects of spatial data are explicitly taken into account into a mathematical model; so then, it is possible to obtain more precise estimates of crop yield response to PA methodologies and technologies, draw valid inferences, and determine potential benefits of robust statistical supported recommendations.

Keywords: Site-specific crop management, precision agriculture, spatial econometrics, geographical information technologies

1. INTRODUCCIÓN

La Agricultura de Precisión (AP) nace a principios de la década de los 90 como una respuesta filosófica y tecnológica al manejo ineficiente de los recursos naturales y económicos por parte de la mayoría de los productores y proveedores de insumos dentro del sector agropecuario y cuyo efecto sobre la agricultura se traduce en el establecimiento de una actividad cada vez menos sostenible y poco amigable con el ecosistema local.

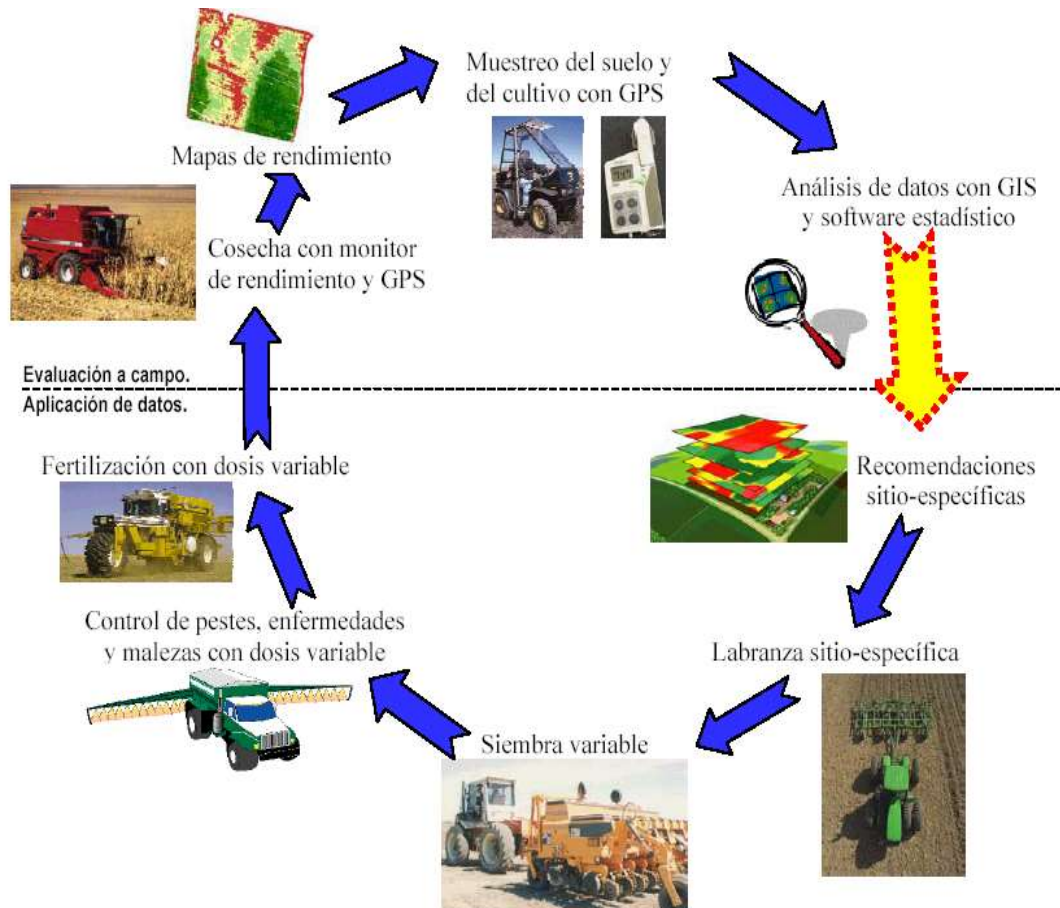


Figura 1. Pasos involucrados en la AP y la Brecha de Información
Fuente: Bongiovanni R. (2002)

En pocas palabras, y tal como se muestra en la Figura 1, la AP implica la adopción de un manejo sitio-específico del cultivo (SSCM, del inglés: Site-Specific Crop Management) que consiste en conocer y determinar las verdaderas necesidades de una plantación en cuanto a insumos, labores agrícolas, microclima y condiciones edáficas, para cada zona dentro de cada unidad de manejo (parcela, cantero, lote) que se pueda considerar homogénea en cuanto a las principales características que afectan al rendimiento del cultivo. Este principio es desde hace ya muchos años considerado como fundamental y necesario para optimizar la aplicación de insumos (fertilizantes, herbicidas, semilla, etc.) y obtener mayores y mejores rendimientos. Sin embargo, su aplicación práctica se consideró desde un inicio poco apropiada debido a los altos costos en mano de obra que requería en contraste al considerable incremento en la producción ocasionado por el manejo a gran escala de los campos mediante la utilización de maquinaria especializada que automatizaba la mayoría de las principales labores agrícolas y que, hasta la actualidad, permiten trabajar muchas más hectáreas en cada vez menos tiempo.

Con la llegada de las tecnologías de información geográfica representadas por los Sistemas de Información Geográficos (GIS, del inglés: Geographical Information Systems), Sistemas de Posicionamiento Global por Satélite (GPS, del inglés: Global Positioning Systems), monitores de rendimientos y maquinaria agrícola que permite la aplicación variable de insumos (VRA, del inglés: Variable Rate Application) o manejo diferencial, el concepto de un manejo sitio-específico vuelve a tener acogida y su rentabilidad comienza a ser analizada.

Son dos los problemas fundamentales a tratarse cuando se requiere demostrar la factibilidad de un proyecto de AP: 1) Determinar el potencial de aplicación de la AP en un determinado cultivo [Mc Bratney *et al*, 2000]; y, 2) La existencia de una brecha de información (flecha amarilla en la Figura 1) entre el análisis de información con referencia espacial y las recomendaciones sitio-específicas de insumos y labores agrícolas [Bongiovanni *et al*, 2002]. Las preguntas críticas que un productor debe responder para poder encontrar una solución eficaz a estos problemas son tres:

- 1.¿Cuánto varían las características medidas del suelo y cultivo?**
- 2.¿Cuánto afectan estas variaciones al rendimiento del cultivo?**
- 3.¿Puede el productor obtener suficiente información y la tecnología apropiada para manejar esta variabilidad con provecho económico?**

Primero, el productor debe determinar cuánta variabilidad existe en las características del suelo y del cultivo. El próximo y más importante paso es encontrar las relaciones *causa-efecto* entre las variables medidas y obtener el resultado en términos de rendimiento y calidad. Esto significa saber cuánto puede producir cada área del campo y que es lo que esta evitando que un área este rindiendo su máximo potencial. Finalmente, si un productor puede averiguar qué está causando la diferencia en rendimiento o calidad, debe determinar si las herramientas y recursos que posee pueden resolver el problema.

El objetivo de esta investigación es disminuir la brecha de información antes mencionada mediante técnicas estadísticas que permitan incorporar los efectos espaciales de los datos en un modelo estocástico que represente una relación causa-efecto entre la producción de un cultivo y sus factores determinantes con la finalidad de establecer un diagnóstico de la situación del cultivo en un ciclo determinado. En particular, se expone el caso del análisis de información proveniente de unidades de manejo de precisión en un cultivo de caña de azúcar en el Ingenio San Carlos (ISC).

Los objetivos específicos son:

1. Determinar cuánto varían las características del suelo y del cultivo: producción y factores determinantes
2. Estimar dos funciones de la respuesta sitio-específica del cultivo a sus factores de producción: la primera corresponde a un modelo econométrico espacial estimado mediante el método de Máxima Verosimilitud (MV), y la segunda corresponde a un modelo econométrico tradicional estimado con el método de Mínimos Cuadrados Ordinarios (MCO).
3. Determinar cuál es la especificación estadísticamente más robusta y correspondiente método de estimación mediante pruebas de diagnóstico y medidas de ajuste.
4. Establecer recomendaciones de manejo agrícola a partir de un diagnóstico de la situación del cultivo sustentado por a especificación en el objetivo 3.
5. Contrastar supuestos de la teoría agronómica para tener un mejor entendimiento del sistema natural bajo estudio.

Desde el punto de vista del análisis estadístico, se busca encontrar los mejores estimadores para los parámetros del modelo econométrico de la producción de caña de azúcar. Dentro del contexto agronómico, el interés está en establecer una metodología que permita sustentar recomendaciones de manejo y realizar inferencias sobre el sistema productivo utilizando información empírica y de mejor calidad contenida en los datos de AP.

2. MATERIALES Y MÉTODOS

Uno de los problemas de la estadística tradicional que se utiliza en el análisis de parcelas experimentales y de ensayos en campo es que se asume que las observaciones son independientes entre ellas. Para el caso particular de los datos de AP, este supuesto de independencia ya no es válido. Las observaciones recogidas por unidades de manejo de precisión o por monitores de rendimiento están claramente relacionadas con las observaciones vecinas a través de cierta dependencia de tipo espacial.

Esta relación de dependencia entre los datos forma parte de los *efectos espaciales* presentes en la información con referencia espacial, los que hacen que ciertas técnicas estadísticas de uso frecuente se vuelvan inapropiadas, tal es el caso del análisis de varianza (ANOVA) proveniente de un diseño experimental de bloques completos aleatorizados para evaluar tratamientos de insumos y los modelos de respuesta del cultivo en base a estimadores de Mínimos Cuadrados Ordinarios (MCO). De aquí que, una de las posibles explicaciones a la brecha de información es que la estructura espacial de los datos no se tiene debidamente en cuenta, lo que afecta tanto a las funciones de respuesta sitio-específica del cultivo como a los análisis de rentabilidad de la AP [Bongiovanni *et al*, 2002].

En este trabajo se estima estadísticamente una función para la respuesta de un cultivo a sus factores de producción a través de la utilización de la econometría tradicional y la econometría espacial [Anselin, 1988] (modelo de regresión múltiple espacial). La razón principal para hacer uso de la econometría espacial es que toma en cuenta de manera explícita tanto la dependencia espacial como la heterogeneidad espacial, los denominados efectos espaciales de los datos georeferenciados.

La dependencia espacial o autocorrelación espacial entre las observaciones de una muestra de tamaño n en el espacio implica que cada observación sobre una variable y en $i \in S$ (donde S es el conjunto que contiene todas las unidades espaciales de observación) está relacionada formalmente a través de una función f con las magnitudes para la variable en otras unidades espaciales del sistema (Expresión 1); es decir, lo que se observa en un punto en el espacio es determinado (en parte) por lo que sucede en otros puntos, por lo que, la posición relativa de unas observaciones con respecto a otras se considera importante y es incluida explícitamente en el análisis.

$$y_i = f(y_1, y_2, y_3, \dots, y_n) \quad (1)$$

La heterogeneidad espacial es consecuencia de la falta de uniformidad de los efectos espaciales. Algunos factores, tales como la existencia de regiones distinguibles, jerarquías de lugares u otras características específicas de algunas zonas, etc., dan los argumentos necesarios para modelar estrategias que tomen en consideración las condiciones particulares de cada sitio específico (o unidad espacial). En el trabajo econométrico, esta tarea puede llevarse a cabo explícitamente mediante varias formas de cambio estructural (Expresión 2). Específicamente en esta investigación, esta tarea se lleva a cabo realizando pruebas de diagnóstico y estimando modelos que incorporen el cambio estructural en forma de regímenes espaciales (diferentes

coeficientes en subconjuntos espaciales del conjunto de datos) por textura de suelos, sectores administrativos, variedades y tercios de cultivo.

$$y_i = f_i(x_i, \beta_i, \varepsilon_i) \quad (2)$$

El modelo econométrico puede especificarse como una función de tres grupos de variables [Bullock y Bullock, 2000]: a) los insumos aplicados, b) las características de cada sitio que escapan al manejo, y c) el clima que tampoco puede ser manejado. Las características de cada sitio representan un grupo de variables que tienden a ser omitidas de los ensayos agronómicos de respuesta. Además, como estas variables tienden a mostrar una distribución espacial, si se las omite de los modelos estadísticos se produce un sesgo, porque los términos del error están correlacionados espacialmente.

La manera de incorporar formalmente la estructura de dependencia espacial en el modelo es asumiendo una interacción entre objetos espaciales discretos (polígonos o cuadrículas). Esto requiere la especificación de un proceso estocástico espacial con una estructura de observaciones contiguas expresada por medio de una matriz de conectividad espacial (Figura 2) que recoge la noción de una contigüidad binaria entre unidades espaciales, es decir, se expresa una estructura de vecinos cercanos mediante valores discretos, 0 ó 1. Para el caso de las unidades de manejo de precisión, si dos lotes tienen un borde en común de longitud diferente de cero, éstos se consideran vecinos cercanos y el valor de 1 es asignado. Al final, la matriz es estandarizada de tal manera que la suma de las filas sea igual a uno y se pueda interpretar a cada elemento del producto yW como el promedio de las observaciones vecinas para cada observación en el vector y .

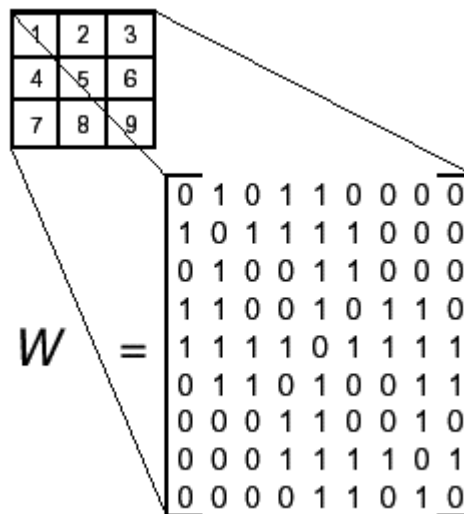


Figura 2. Arreglo de nueve celdas expresado en una matriz de contigüidad espacial W .
Fuente: Bongiovanni *et al* (2002)

Dentro de los modelos de regresión espacial pueden distinguirse algunas especificaciones, de acuerdo a la forma de dependencia espacial que exista entre las observaciones. Partiendo de la especificación más general (Expresiones 3-6) y manteniendo algunas restricciones sobre los parámetros, pueden obtenerse los siguientes modelos [LeSage, 1999]: 1) SAC o especificación general sin heteroscedasticidad con $\Omega = \sigma^2 I$; 2) SAR o modelo de “ponderación espacial” para el que $W_2 = 0$ y $\Omega = \sigma^2 I$ (sin heteroscedasticidad) y donde la variable dependiente esta

correlacionada espacialmente consigo misma; y, 3) SEM o modelo “de error espacial” donde $W_1 = 0$ y $\Omega = \sigma^2 I$; es decir, la variable dependiente no está correlacionada por sí misma sino que es el error estocástico el que está correlacionado espacialmente.

$$y = \rho W_1 + X\beta + u \quad (3)$$

$$u = \lambda W_2 u + e \quad (4)$$

$$e \sim N(0, \Omega) \quad (5)$$

$$\Omega_{ii} = h_i(z'\alpha) \quad (6)$$

Específicamente, para nuestro modelo de producción tenemos:

$$\text{Producción} = \rho W(\text{Producción}) + \beta_0 + \beta_1 \text{Factor 1} + \beta_2 \text{Factor 2} + \dots + \beta_k \text{Factor } k + u$$

$$u = \lambda W u + e$$

$$e \sim N_n(0_n, \sigma^2 I_n)$$

En la especificación (3-6), y es un vector con los valores de la producción para cada lote de la muestra, β es un vector de dimensión $k \times 1$ de parámetros asociados con los factores de producción representados por las variables exógenas (no aleatorias) X (matriz $n \times k$), ρ es el coeficiente de rezago espacial para la variable dependiente y λ es el coeficiente para la estructura autorregresiva espacial del error u .

El componente del error estocástico e está normalmente distribuido con una matriz general de varianzas-covarianzas Ω . Los elementos de la diagonal permiten la presencia de heteroscedasticidad (varianza no constante) como una función de $p + 1$ variables exógenas z , que incluye un término constante, de manera que, para $\alpha = 0$, se tiene $h = \sigma^2$ para todos los términos de error, la situación clásica de homoscedasticidad.

La invalidez de los estimadores de mínimos cuadrados ordinarios (MCO) radica en que en presencia de los efectos espaciales los estimadores de los parámetros del modelo resultan: 1) sesgados¹ e inconsistentes² para el caso de autocorrelación espacial en la variable dependiente que resulta en el coeficiente ρ estadísticamente diferente de cero (significativo) independientemente del coeficiente λ ; o, 2) ineficientes³ para el caso en el que exista únicamente autocorrelación espacial del error; es decir, si el coeficiente λ es estadísticamente significativo con el coeficiente ρ estadísticamente no significativo.

Para solucionar este hecho, se ha centrado la atención en los estimadores de Máxima Verosimilitud (MV) como una alternativa y se ha tomado ventaja de sus atractivas propiedades asintóticas o de muestras grandes [Johnston y Dinardo, 1997]: 1) consistencia, es decir, converge

¹ Un estimador es sesgado cuando su valor esperado no es igual al parámetro que se desea estimar.

² Un estimador es inconsistente cuando no converge en probabilidad al parámetro que se desea estimar.

³ Un estimador es ineficiente cuando no es de mínima varianza.

en probabilidad a los parámetros que se quiere estimar; 2) normalidad asintótica con media igual al vector de parámetros desconocidos; 3) eficiencia asintótica, es decir, de varianza mínima entre todos los estimadores asintóticos; y, 4) invarianza, es decir, si $\hat{\theta}$ es el estimador de MV de θ y $g(\theta)$ es una función continua de θ , entonces $g(\hat{\theta})$ es el estimador de MV de $g(\theta)$.

Los estimadores MV resultan de la maximización de la función log-verosimilitud, la cual representa el logaritmo de la función de densidad probabilística de los valores de la muestra, con respecto a los parámetros desconocidos del modelo; es decir, estamos maximizando la probabilidad de haber obtenido la muestra con respecto a los valores de los parámetros del modelo. Se expresa formalmente a la maximización de la función log-verosimilitud del modelo SAC como:

$$\text{Max}(\theta) \quad L = -\frac{n}{2} \ln(\pi) - \frac{1}{2} \ln|\Omega| + \ln|B| + \ln|A| - \frac{1}{2} e'e$$

donde:

$$\begin{aligned} e'e &= (Ay - X\beta)' B' \Omega^{-1} B (Ay - X\beta) \\ A &= I - \rho W \\ B &= I - \lambda W \\ \Omega &= \sigma^2 I \\ y &= \text{TMC/ha.} \\ \theta &= [\beta', \rho, \lambda] \end{aligned}$$

Las pruebas de hipótesis o “tests” más conocidas basadas en el principio de MV son [Johnston y Dinardo, 1997]: 1) Wald (W); 2) Razón de Verosimilitud (LR, del inglés: Likelihood Ratio); y, 3) Multiplicador de Lagrange (LM, del inglés: Lagrange Multiplier). En términos generales, las tres pruebas asintóticas se basan en las diferentes medidas de distancia entre una estimación sin restricciones y una estimación que satisface las condiciones impuestas por la hipótesis que se desea probar (hipótesis nula), es decir, el estimador restringido. Dentro de este contexto, una prueba para la presencia de autocorrelación espacial en los residuos de un modelo tradicional puede ser formulada de la siguiente manera:

$$\begin{aligned} H_0 &: \lambda = 0 \\ H_1 &: \lambda \neq 0 \end{aligned}$$

Formalmente, la función de log-verosimilitud para el modelo con error autorregresivo espacial es:

$$\begin{aligned} L &= -\left(\frac{n}{2}\right) \ln(\pi) - \left(\frac{n}{2}\right) \ln(\sigma^2) + \ln|I - \lambda W_2| \\ &\quad - \left(\frac{1}{2}\right) \sigma^{-2} (y - X\beta)' (I - \lambda W_2)' (I - \lambda W_2) (y - X\beta) \end{aligned} \quad (7)$$

Bajo la hipótesis nula, es decir, con $\lambda = 0$ y $(I - \lambda W_2) = I$, tenemos la función de log-verosimilitud usual para un modelo de regresión lineal tradicional:

$$L_0 = -\left(\frac{n}{2}\right)\ln(\pi) - \left(\frac{n}{2}\right)\ln(\sigma^2) - \left(\frac{1}{2}\right)\ln(\sigma^2) - \left(\frac{1}{2}\right)\sigma^{-2}(y - X\beta)'(y - X\beta) \quad (8)$$

La prueba LR para la autocorrelación espacial en los residuos esta basada en la diferencia entre las expresiones 7 y 8. Cuando los coeficientes en la función de log-verosimilitud se reemplazan por sus estimadores de MV se tiene la siguiente expresión simplificada:

$$LR = n[\ln(\sigma_0^2) - \ln(\sigma_1^2)] + 2\ln|I - \lambda W| \sim \chi^2(1)$$

donde σ_0^2 es el estimador de la varianza del error para el modelo bajo la hipótesis nula y σ_1^2 es el estimador de la varianza del error para el modelo espacial. El estadístico de prueba está asintóticamente distribuido como una χ^2 con un grado de libertad (correspondiente a una restricción).

La validación estadística del modelo econométrico y la comparación entre especificaciones, que para el caso resultan ser un modelo MCO “no espacial” y otro MV “espacial”, debe basarse en los denominados Criterios de Información [3], que se constituyen en una medida de qué tan cerca se encuentra el modelo postulado del verdadero modelo. El criterio más utilizado es el de Akaike (AIC, del inglés: Akaike Information Criterion) que maximiza la entropía del modelo formulado corrigiendo el valor del logaritmo de la función de verosimilitud evaluada en los estimadores MV (maximizada) por un factor que refleja el número de parámetros utilizados; entre algunos modelos candidatos, se prefiere aquel con un menor valor en la medida de AIC.

2. DATOS

Los datos utilizados en este trabajo fueron obtenidos en el ISC ubicado en el cantón Marcelino Maridueña, provincia del Guayas, Ecuador (80.00° O y 01.81° S), y corresponden al ciclo de cultivo de caña de azúcar 2002-2003. Las precipitaciones fueron de 892 mm, el total de horas de sol fue de 629,8, la temperatura promedio fue de 25,28°C con un máximo de 28,98°C y un mínimo de 21,54°C. Es una plantación de aproximadamente 16.500 hectáreas con un ciclo de cultivo de 12 meses. La cosecha o zafra comienza a mediados del mes de junio y termina a finales de diciembre con la llegada del invierno.

Tabla 1. Estadística descriptiva de los parámetros del análisis de suelos

	CE (mmho/cm)	pH	MO (%)	P (ppm)	K (cmol/L)	Ca (cmol/L)	Mg (cmol/L)
Lotes	871	871	871	871	871	871	871
Media	72,0103	6,2721	1,8236	15,5029	,4015	16,8093	8,2450
Mediana	63,0000	6,2600	1,8200	12,0000	,3200	16,1600	7,0000
Desv. Est.	36,40454	,42354	,45456	12,71228	,25132	5,20702	4,45470
Mínimo	25,00	4,77	,35	1,00	,06	6,76	1,08
Máximo	334,00	8,17	3,55	108,00	1,54	32,16	27,84
Percent.							
25	50,0000	5,9800	1,5200	7,0000	,2300	12,6000	4,9600
50	63,0000	6,2600	1,8200	12,0000	,3200	16,1600	7,0000
75	83,0000	6,5300	2,1400	19,0000	,4900	20,5200	10,6400

Tabla 2. Estadística descriptiva de la producción y algunos de sus factores

	Producción (TMC/ha.)	Has	EDAD (años)	Edad a la cosecha (meses)	Número de riegos	
Lotes	871	871	871	871	871	
Media	75,2013	12,7046	3,8634	12,0623	7,2457	
Mediana	76,5900	11,8900	4,0000	12,1200	7,0000	
Desv. Est.	18,12659	6,57123	2,19161	,65321	2,70066	
Mínimo	20,89	,98	1,00	10,13	2,00	
Máximo	140,24	55,71	13,00	16,11	16,00	
Percent.	25	62,6067	7,6800	2,0000	12,0100	5,0000
	50	76,5900	11,8900	4,0000	12,1200	7,0000
	75	87,9654	17,2400	5,0000	12,2400	9,0000

El ISC está dividido con fines de manejo administrativo y agrícola en 6 sectores de aproximadamente 3000 ha. cada uno, 14 secciones y 347 canteros. La división por sectores es la siguiente:

- SECTOR 1: Parte de la sección 1 y 6, y la sección 5.
- SECTOR 2: Comprende la sección 2 y 20
- SECTOR 3: Sección 3, 4 y 17
- SECTOR 4: Sección 9 y 10
- SECTOR 5: Sección 8, 11, 18, 19
- SECTOR 6: Parte de la sección 1 y 6, y la sección 7.

Hasta finales del 2002, el cantero era considerado como la mínima unidad de manejo; es decir, un área dentro de la cual la aplicación de tratamientos agrícolas se realizaba de manera uniforme bajo el supuesto de obtener una respuesta uniforme del cultivo. Sin embargo, se conoce que dentro del mismo cantero existe variabilidad en la producción de caña de azúcar debido, muy probablemente, a su gran extensión (se encuentran canteros de 3 hasta 138 ha., con 48 ha. de promedio) y al efecto de las diferentes condiciones en cuanto a textura de suelo, niveles de fertilidad, drenaje, variedad de caña, entre otros factores.

Con la implementación de la AP en el ISC, se identificó la necesidad de disminuir el área de la unidad de manejo establecida, es decir, subdividir los canteros en lotes. Lotizar, dentro del contexto de AP, tiene como objetivo establecer unidades de manejo de menor extensión lo más homogéneas posible en cuanto a las condiciones presentes de los factores más importantes que inciden sobre la producción, para luego, proceder a recabar la mayor cantidad de información sobre el cultivo por cada nueva unidad de manejo o lote, de forma que, cuando se tome una decisión de manejo, esta tenga una mayor probabilidad de ser la correcta en comparación a la de un cantero.

El criterio principal para el establecimiento de los lotes fue dividir los canteros por tipos de suelo, para lo cual, se procedió a verificar la variabilidad de la textura del suelo cantero por cantero apoyándose en un mapa de suelos establecido anteriormente. Como resultado, se diseñaron 1309 lotes con un promedio de 13 ha. (en un rango de 2 a 20 ha.). Muy pocos canteros quedaron con lotes mayores a 20 ha., los cuales, en su siguiente renovación serán rediseñados para que sus divisiones tengan áreas menores. Los lotes se identifican con un código de seis dígitos y toda la información de cosecha (producción, camiones, zafreros, etc.) se recolecta y registra al nivel de lotes y no de canteros.

Para el monitoreo tanto de la producción como de sus factores se adquirieron tecnologías de información geográfica, específicamente, unidades GPS para referenciar espacialmente los límites de cada lote y un GIS como soporte a la toma de decisiones y control de labores agrícolas (fertilización, riego, aplicación de herbicidas, etc.).

La información inmediatamente disponible en las capas del GIS para las nuevas unidades de manejo de precisión corresponde a:

1. Producción: Toneladas métricas de caña de azúcar por hectárea (TMC/ha.)
2. Análisis de suelos: Textura de suelos, parámetros del análisis de suelos (P, K, Ca, pH, Mg, materia orgánica (MO), Conductividad Eléctrica (CE))
3. Sector donde se ubica el lote.
4. Tercio de inicio de cultivo: a) julio-agosto (primer tercio); b) septiembre-octubre (segundo tercio); y, c) noviembre-diciembre (tercer tercio).
5. Variedad de caña de azúcar.
6. Edad del cultivo (años).
7. Edad a la cosecha (meses).
8. Número de riegos

El análisis estadístico exploratorio de cada una de las variables que representan tanto a la producción como a sus factores se realizó sobre una muestra de 871 lotes y los resultados se muestran en las Tablas 1, 2 y 3.

Tabla 3. Estadística descriptiva de los factores cualitativos

		LOTES	%
SECTOR	1	124	14,2%
	2	134	15,4%
	3	140	16,1%
	4	197	22,6%
	5	164	18,8%
	6	112	12,9%
TERCIO	TERCIO 1	310	35,6%
	TERCIO 2	308	35,4%
	TERCIO 3	253	29,0%
VARIEDAD	B-74132	14	1,6%
	BJ-7046	54	6,2%
	CC-8592	64	7,3%
	CR-74250	176	20,2%
	MISCELANEA	27	3,1%
	NCO-310	6	,7%
	PR-671070	86	9,9%
	RAGNAR	444	50,9%
TEXTURA	Arcillo Limoso	9	1,0%
	ARCILLOSO	198	22,7%
	Areno Limoso	15	1,7%
	ARENOSO	20	2,3%
	FRANCO	155	17,8%
	Franco Arcilloso	271	31,1%
	Franco Arcilloso Arenoso	28	3,2%
	Franco Arcilloso Limoso	14	1,6%
	Franco Arenoso	139	15,9%
	Franco Limoso	22	2,5%

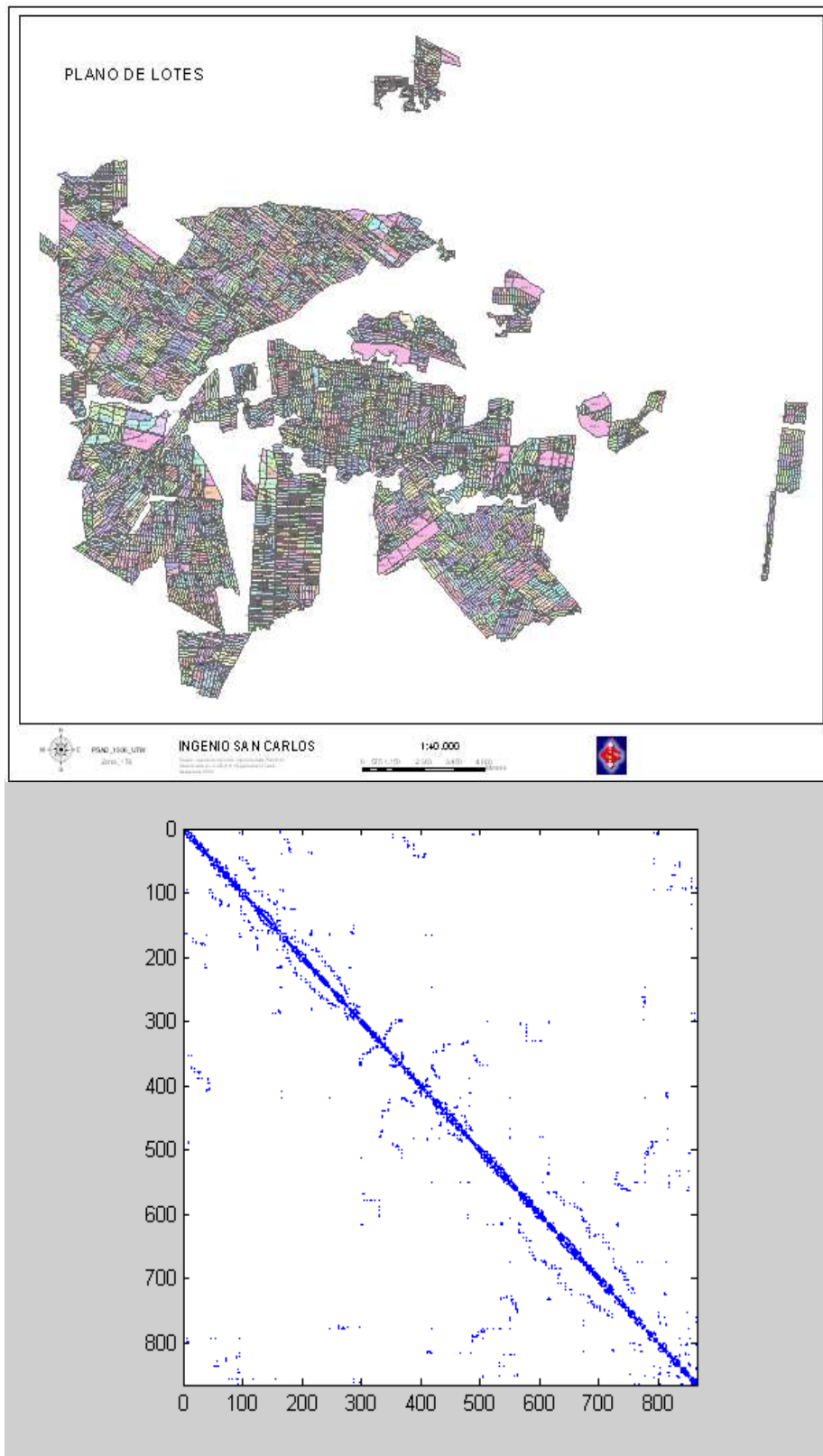


Figura 3. Plano de lotes del ISC y su matriz de contigüidad espacial W
Fuente: Dep. Campo, ISC

3. RESULTADOS

Dentro de los resultados obtenidos, uno de los más importantes es la diferencia entre un modelo tradicional MCO que incluye los regímenes por sector, tercio, variedad y textura y un modelo “espacial” MV (SAC) que incluye un término autorregresivo tanto para la variable dependiente (producción) como para el término de error y una matriz de contigüidad espacial W que se muestra en la Figura 3. Los resultados de la estimación se muestran en la Tabla 3.

Tabla 3. Modelos estimados y medidas de ajuste

VARIABLES	MCO		MV		Estadístico de colinealidad	
	COEF	SIG	COEF	SIG	VIF	
Constante	33,911	,001	50,582085	,000000		
S3: sector 3	8,453	,000	9,103924	,000009	1,738	
RIEGOS	,553	,012	0,749208	,003186	2,787	
T1: primer tercio	22,548	,000	22,428621	,000000	3,961	
T2: segundo tercio	15,213	,000	14,195213	,000000	2,131	
T1*OTV	-5,652	,009	-6,118555	,002978	1,173	
T3*PR-671070	6,867	,046	7,531252	,022616	1,274	
T1*FARC	3,062	,055	3,589242	,020323	2,511	
F*CC-8592	8,291	,016	7,400152	,020168	1,479	
FARE*CC-8592	12,202	,028	10,250813	,028686	1,110	
FARE*PR-671070	12,756	,000	8,222838	,002335	1,262	
Edad cultivo	-1,913	,000	-1,999399	,000000	1,868	
CC-8592	2,916	,111	4,465887	,010601	1,802	
CR-74250	6,553	,000	9,217573	,000000	1,729	
PR-671070	-8,829	,000	-6,329794	,000018	1,610	
F	5,708	,000	3,459848	,001200	1,315	
FARC	4,414	,000	1,900156	,074510	1,965	
K	4,157	,012	3,382351	,046131	1,366	
pH	-,714	,501	-1,884619	,006517	1,600	
CE	-,030	,009	-0,026363	,010436	1,364	
Edad a la cosecha	2,349	,000	2,114313	,000029	1,324	
Ca	,521	,000	0,363828	,002943	2,921	
Mg	-,732	,000	-0,426553	,002297	3,219	
Rho	-	-	-0,090999	,000004	-	
Lamda	-	-	0,611000	,000000	-	
Medidas de ajuste			MCO	SIG	MV	SIG
R^2 Ajustado			0,674		0,7427	
F-test			79,643	0,000		
Log-verosimilitud			-3271,090		-2404,6969	
Criterio de Información de Akaike (AIC)			6588,18		6441,252	

Todos los coeficientes del modelo MV son significativos al 0.05 excepto FARC (al 0.1 sig.)

Incluidas entre las pruebas de diagnóstico, se realizaron las pruebas de heteroscedasticidad de White para probar la hipótesis de varianza constante del error, Kolmogorov-Smirnov para probar normalidad del error del modelo MV (SAC), pruebas basadas en el principio MV (Wald, LR y LM) para probar la hipótesis de no autocorrelación espacial del error en el modelo MCO y, por último, una prueba de no autocorrelación espacial del error en un modelo SAR. Los resultados se muestran en la Tabla 5.

Tabla 5. Pruebas de diagnóstico

Pruebas de diagnóstico	g.l.	valor	Sig.
Heteroscedasticidad White (términos no cruzados MCO)	30	45,48967	0,034735
Heteroscedasticidad White (términos cruzados MCO)	275	195,8999	0,528846
LR (error MCO)	1	129,2769	0,000000
Wald (error MCO)	1	1914,697	0,000000
LM (error MCO)	1	137,5619	0,000000
LMSAR(error SAR)	1	391,8294	0,000000
Kiologorov-Smorov Z (error est. SAC)		1,355	0,051000

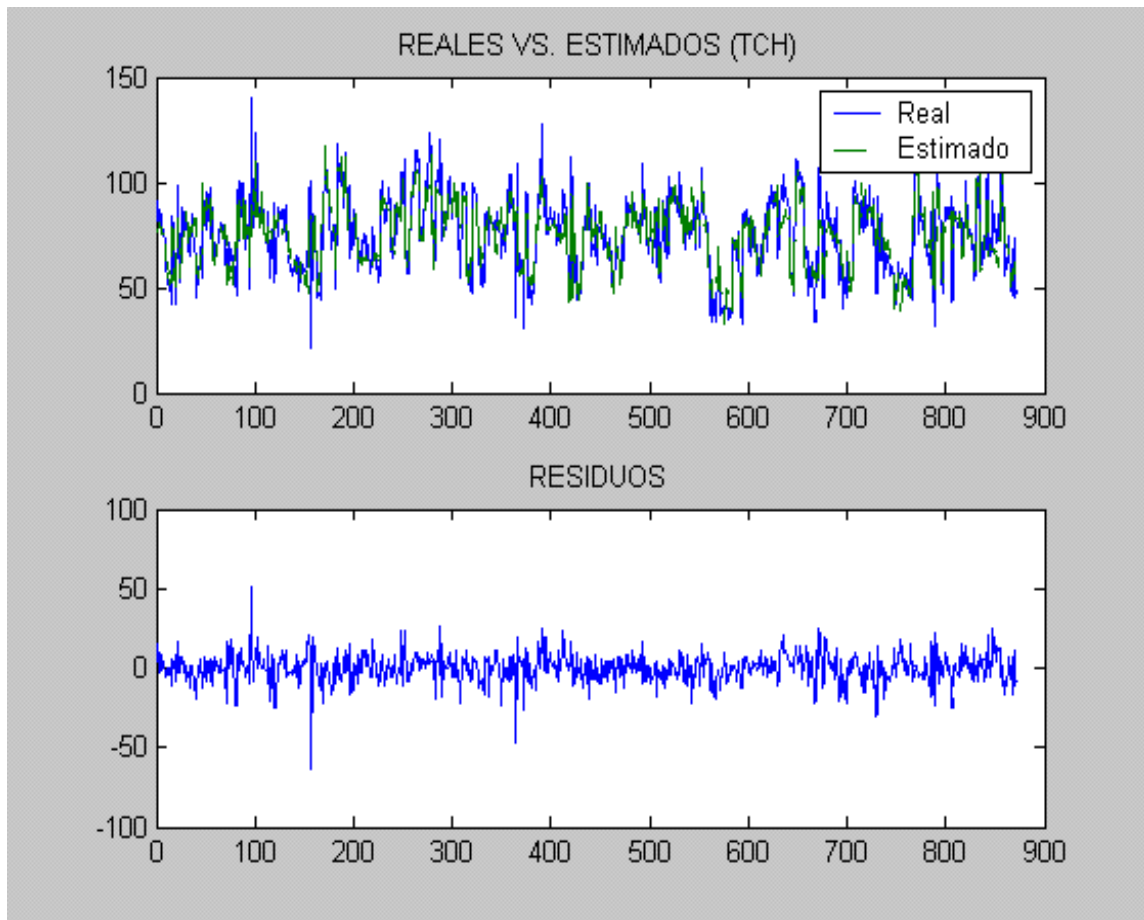


Figura 4. Modelo espacial MV: Valores reales vs. estimados y residuos

De los resultados de las pruebas realizadas, se adoptaron los supuestos de no heteroscedasticidad, errores autocorrelacionados espacialmente en el modelo MCO, errores autocorrelacionados espacialmente en el modelo SAR (que incluye únicamente el rezago espacial de la variable dependiente) y normalidad del error en el modelo espacial MV (SAC). Este último supuesto es importante pues nos permite realizar inferencias sobre la significancia estadística de cada uno de los parámetros del modelo.

La presencia de un rezago estadísticamente significativo a los niveles 0.01, 0.05 y 0.1 en la variable dependiente y el error estocástico hace que los estimadores MCO sean sesgados e ineficientes, por lo que, el modelo MV resulta ser el más adecuado. Además, la medida de AIC (medida de entropía) es menor en el modelo espacial MV lo que indica un mejor ajuste en comparación al modelo MCO. La Figura 4 muestra la proximidad de los valores estimados (en verde) por el modelo y los valores reales (en azul).

Dentro de la interpretación agronómica y mediante la evidencia empírica que aporta el modelo MV, se pudieron evaluar algunos supuestos, dentro de los cuales, los de mayor interés fueron los siguientes:

1. El sector tres presenta una mayor producción en contraste a los demás sectores.
2. El primer y segundo tercio de cultivo presentan las mejores producciones debido a las interacciones climáticas-fisiológicas más favorables en comparación al tercer tercio.
3. El aumento en la edad del cultivo incide negativamente en la producción.
4. La variedad CR-74250 es la más productiva.
5. La textura franco del suelo es la que más favorece a la producción.

El sector 3 está representado en el modelo por la variable ficticia S3 que toma el valor de 1 si el lote pertenece al sector tres y 0 si no. El valor del coeficiente para la variable indica que el sector aporta aproximadamente 9 TMC/ha. más que los otros sectores. Este resultado era previsto debido a que desde hace 20 años el sector recibe un número considerable de riegos con agua residual proveniente de la fábrica donde se procesa la caña y vinaza de la destilería de alcohol. Estos subproductos son ricos en nutrientes como fósforo (P) y potasio (K) y mejoran las características químicas del suelo. Cuando se inició el riego con esta agua, el sector 3 presentaba los valores de productividad más bajos, después de algunos años de riego la producción comenzó a incrementarse cada año hasta convertirse en el sector con mayor tonelaje. Estos resultados demuestran los beneficios de la incorporación de materia orgánica al suelo.

El primer tercio de cultivo está representado por la variable ficticia T1 y los resultados indican que este tercio aporta aproximadamente 22,43 TMC/ha. más que el tercer tercio. El segundo tercio está representado por la variable ficticia T2 y el periodo de cultivo aporta aproximadamente 14,19 TMC/ha. más que el tercero. Este resultado está de acuerdo con las expectativas, debido a que el primer y segundo tercios presentan las condiciones climáticas más favorables para el periodo de desarrollo fisiológico correspondiente al rápido crecimiento de la caña (7 meses).

La edad del cultivo representada por la variable cuantitativa Edad de cultivo presenta un coeficiente negativo. Se estima que por cada año transcurrido la producción decrece en aproximadamente 2 TMC/ha. Este resultado es esperado debido a que se pierde densidad en la plantación por las distintas labores agrícolas realizadas año tras año. Lo importante es cuantificar estas pérdidas para poder expresarlas en costos.

El coeficiente de la variable Edad a la cosecha puede interpretarse que por cada mes que se deja rezagada la caña, es decir, por cada mes después de los 12 meses del ciclo que se retrase la cosecha, la producción aumenta en promedio 2,11 TMC/ha. Este resultado es importante en la medida en que permite investigar sobre la práctica discutida de rezagar caña para obtener mejores producciones.

La variable RIEGOS se incluyó en el modelo con el fin de determinar en parte la influencia del suministro de agua por medio del riego al cultivo. Esta variable no considera todas las formas de suministro de agua ni el momento en que se lo hace, por lo tanto, no se la debe considerar un indicador absoluto de la efectividad del suministro de agua. Sin embargo, se la considera útil para mostrar si existe alguna tendencia en la producción con respecto al número de riegos aplicados. El coeficiente para la variable indica que por cada riego que dejó de aplicarse se perdió el potencial de producir en promedio 0,75 TMC/ha. Este resultado permite cuantificar las pérdidas en producción. Esta situación podría estar indicando que están faltando riegos durante el periodo de rápido crecimiento de la caña (3 – 8 meses) que es cuando la planta necesita el agua para producir biomasa.

Dentro de las variedades representadas por variables ficticias, los resultados indican que la variedad CR-74250 permite obtener aproximadamente 9,23 TMC/ha. más que la variedad predominante RAGNAR (que no se muestra en el modelo pues es tomada como la variedad base para hacer comparaciones). Le sigue la variedad CC-8592 con aproximadamente 4,46 TMC/ha. sobre la RAGNAR. Se puede considerar a la variedad PR-671070 como la de menor producción pues aporta con aproximadamente 6 TMC/ha. por debajo de la variedad Ragnar, y, consecuentemente, del resto de variedades.

Sin embargo, debemos considerar las interacciones de las variedades con las texturas del suelo pues el modelo permite concluir que en la textura franco arenosa, la variedad PR-671070 ya no es la de menor producción. Dentro de las recomendaciones agrícolas se debería sembrar esta variedad en texturas franco arenosas. Un resultado similar puede observarse con la variedad CC-8592.

Analizando las texturas, representadas por variables ficticias, todas resultan ser estadísticamente significativas al 0.05, 0.01 y 0.1, a excepción de la variedad creada OTRT estadísticamente significativa al 0.05 y 0.1. Esta última variable agrupa a todas las variedades que se encuentran presentes en un número reducido de lotes por lo que no tiene una clara interpretación. La textura franco (F) resulta la más favorable para el crecimiento de la caña, seguida de la franco-arcillosa (FARC) y la franco-arenosa (FARE), todas éstas se comparan a la textura arcillosa no incluida explícitamente en el modelo por considerársela como base para realizar comparaciones. Estos resultados, además de estar acorde con los conocimientos agronómicos básicos, permitieron evaluar como apropiado el criterio de lotización por texturas ya que, de acuerdo al modelo y al presentarse las variables de textura significativas, queda demostrado que esta característica del suelo explica significativamente la variabilidad en la producción.

4. CONCLUSIONES

La metodología descrita en esta investigación es un paso clave hacia la interpretación de la información proveniente de unidades de manejo de precisión. Las técnicas tradicionales, como el caso específico tratado en este análisis de la utilización de estimadores MCO, nos conducen hacia estimadores sesgados, inconsistentes e ineficientes, y por ende, hacia conclusiones incorrectas. Según las pruebas de diagnóstico y las medidas de ajuste aplicadas (AIC menor) utilizadas, la especificación espacial SAC o modelo MV es la más apropiada para aproximar el sistema natural de la producción de caña de azúcar debido a que adopta supuestos más cercanos a la realidad.

El beneficio fundamental del uso de la econometría espacial es que se toma en cuenta la estructura espacial de los datos de AP y como consecuencia los estimadores de los parámetros al modelo de la producción de caña de azúcar son estadísticamente más robustos. Especificaciones correctas que representen fielmente al sistema producción se traducirán en inferencias válidas y

cualquier tipo de análisis que se realice sobre la base de éstas reflejaran los verdaderos beneficios potenciales de la utilización de AP en un cultivo.

Los problemas de la adopción de la AP, específicamente en cuanto a la brecha de información se refiere, pueden ser superados cuando exista suficiente información de calidad para apoyar decisiones (recomendaciones) de tipo sitio-específico y cuando la información disponible se utilice eficientemente.

Reconocimientos

Los autores desean agradecer la invaluable colaboración de Egbert Spaans, Phd. en suelos y profesor de suelos en la universidad EARTH, Costa Rica; y, Ing. Leonidas Estrada, coordinador de AP, Ingenio San Carlos.

REFERENCIAS

- Anselin, L. (1988) Spatial Econometrics: methods and models. (London, Kluwer Academic Publishers, 1988)
- Bongiovanni, R., Lowenberg-DeBoer J. y Lambert D. (2002), Proceedings of the 6th International Conference on Precision Agriculture, (Bloomington, MN, julio 14-17, 2002) "Spatial Regression, an Alternative Statistical Analysis for Landscape Scale On-farm trials: Case Study of Variable Rate Nitrogen Application in Argentina"
- Bongiovanni, R. (2002) Tercer Taller Internacional de Agricultura de Precisión del Cono Sur de América, (Carlos Paz, Córdoba, Argentina, diciembre 17-19, 2002) "Econometría Espacial: Una herramienta clave para el manejo sitio-específico de insumos"
- Bullock, D. S. y Bullock D. G. (2000) "From Agronomic Research to Farm Management Guidelines: A primer on the Economics of Information and precision Technology." En: [Bongiovanni, 2002]
- Johnston, J. y Dinardo, J.(1997) Econometric Methods. (4ta. Edición, McGraw-Hill, 1997), pp 142-151.
- LeSage, J. P. (1999) The Theory and Practice of Spatial Econometrics (University of Toledo, Department of Economics, febrero 1999), pp.43-97, [http:// www.econ.utoledo.edu](http://www.econ.utoledo.edu).
- McBratney, A.B.,Whelan, B.M.,Taylor, J.A. y Pringle, M.J. (2000), Proceedings of the 5th International Conference on Precision Agriculture and Other Resource Management, (Bloomington, MN, julio 16-19, 2000) "A Management opportunity index for precision agriculture, por A.B. McBratney, B.M. Whelan, J.A. Taylor y M.J. Pringle."