

Clasificación supervisada de suelos de uso agrícola en la zona central de Córdoba (Argentina): comparación de distintos algoritmos sobre imágenes Landsat

Enrique Willington¹, Miguel Nolasco¹ y Mónica Bocco¹

¹ Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba. Argentina.
enriquewillington@gmail.com, mnqn_88@hotmail.com, mbocco@gmail.com

Resumen. El tipo de cobertura del suelo es un dato de mucha importancia en el ámbito del manejo de los recursos naturales, siendo uno de los factores que cambia más rápidamente también en el sector agrícola. La clasificación de imágenes provenientes de sensores remotos permite monitorear dichos cambios con precisión, en forma periódica y a menores costos que con métodos tradicionales. Las imágenes del satélite Landsat TM son ampliamente utilizadas para realizar dichas clasificaciones. La soja (*Glycine max* (L.) Merrill) y el maíz (*Zea mays* L.) son los cultivos más importantes en Argentina, por área sembrada y réditos económicos producidos. En el presente trabajo se clasificaron dos imágenes Landsat, empleando cuatro métodos de clasificación supervisada, utilizando el software Envi 4.6. El objetivo fue evaluar y comparar los métodos aplicados para el mapeo de la cobertura del suelo, en el área central de Córdoba (Argentina). El desempeño de los clasificadores fue muy bueno, siendo el de máxima probabilidad el que presentó mejores resultados, con estadístico Kappa catalogado como excelente, tanto para la imagen de inicio de la campaña agrícola como para la de finales de de la misma.

Palabras clave: clasificación, cultivos, cobertura, Landsat TM.

1 Introducción

Los datos terrestres de uso y cobertura del suelo están entre los más importantes y universalmente utilizados en el ámbito de la gestión sostenible de recursos y en estudios de indicadores ambientales y de cambio climático, proveyendo información medioambiental clave para muchas aplicaciones científicas. Por esto es cada vez más necesario poseer datos exactos, detallados y periódicos [1-2]. El tipo de cobertura del suelo es una variable trascendente en los ambientes rurales, la variación de la vegetación es muy importante en el cambio global, y ambos afectan los sistemas ecológicos con un impacto ambiental que puede ser tan grande como el asociado con el cambio climático. Los datos de cobertura, en forma de mapa, no están fácilmente disponibles; además, de los elementos representados en los mismos, la cobertura del suelo está entre los que cambian más rápidamente, por lo tanto la capacidad para monitorearla con precisión es importante [3].

La información provista por sensores remotos es importante para estudios relacionados con el uso y ocupación de la superficie del suelo; siendo la clasificación

de imágenes uno de los problemas técnicos más importantes de la teledetección. Entre las ventajas de ésta respecto de otros métodos que obtienen datos a campo se encuentran un menor costo, cobertura de grandes áreas y la posibilidad de repetir la obtención de los datos a lo largo del tiempo [4].

Existen varios tipos de imágenes utilizadas para monitorear la superficie terrestre. Las clasificaciones de cobertura de suelo son, generalmente, hechas a partir de imágenes Landsat TM [5]; en Argentina, Guerschman *et al.* [6] utilizaron datos multitemporales de Landsat TM, de la misma estación de crecimiento, para clasificar tipos de cobertura y Wardlow *et al.* [2] investigaron la aplicabilidad de series temporales de datos de MODIS 250 m para clasificar uso/cobertura del suelo en cultivos agrícolas en las planicies centrales de EEUU. Para Córdoba, Argentina, Bocco *et al.* [7] clasificaron imágenes MODIS, utilizando redes neuronales, según la cobertura del suelo en las categorías: soja, maíz y suelo desnudo.

Los diferentes métodos paramétricos que existen para clasificar el uso o cobertura del suelo pueden agruparse en tres categorías principales: basados en distancia, en probabilidad y en reglas de decisión angular. Los primeros utilizan los valores espectrales medios de las distintas clases, ignorando las varianzas entre ellas; los segundos ofrecen, generalmente, mayor exactitud que los anteriores; incorporando la media y la varianza de los datos utilizados, lo que les proporciona información adicional sobre la cual basar la clasificación. Los terceros utilizan una regla de decisión basada en ángulos espectrales formados entre un espectro de referencia y un píxel no clasificado en un espacio de n dimensiones, donde n es el número de bandas disponibles [8]. Los métodos de clasificación pueden ser supervisados y no supervisados; los primeros son procedimientos empleados para la identificación de áreas espectralmente similares (muestras de entrenamiento) dentro de una imagen, por lo que el usuario preliminarmente reconoce las regiones de interés conocidas en el área de tierra, y el algoritmo elegido extrapola estas características espectrales para otras regiones de la imagen, realizando así la clasificación [9]. La clasificación no supervisada tiene por objetivo agrupar los casos por su similitud espectral relativa, sin toma de muestras a campo [3]. Vizzotto Cattani *et al.* [10] compararon cuatro métodos de clasificación supervisada, analizando el potencial individual de cada clasificador para el mapeo de cultivos agrícolas a partir de imágenes multi-espectrales de los satélites RapidEye.

En Argentina, por superficie sembrada y réditos económicos, tanto para los productores como para el Estado Nacional, el cultivo de soja (*Glycine max* (L.) Merrill) es el más importante. En la campaña 2011/12 se sembraron a nivel nacional 18.670.000 ha y en Córdoba 5.014.000 ha (27 % del total nacional). Por otro lado el cultivo de maíz (*Zea mays* L.) es el segundo en importancia en el país, con un área total sembrada superior a 5.000.000 ha en la misma campaña. Para este cultivo nuestra provincia es el segundo productor con una superficie sembrada en 2011/12 de aproximadamente 1.272.000 ha. [11]

El objetivo de este trabajo fue evaluar y comparar el desempeño de cuatro métodos de clasificación supervisada, en la identificación y mapeo de cobertura de suelo a partir de imágenes satelitales Landsat. Determinando la rutina de clasificación más adecuada para la zona de trabajo.

2 Materiales y métodos

2.1 Área de estudio

El estudio fue realizado en la región central de la provincia de Córdoba, Argentina (departamentos Río Segundo y Santa María, cuya ubicación aproximada es 32° S; 64° W). El área presenta un relieve ligeramente ondulado con una leve pendiente hacia el este, suelos de textura franco limosa clasificados como Haplustoles énticos y típicos.

El clima es clasificado como seco sub-húmedo, con precipitaciones anuales promedio de 800 mm, concentradas en el verano [12]. El método de labranza predominante en la zona es el de siembra directa, siendo la principal actividad productiva la agricultura, predominando los cultivos de verano, soja y maíz. En menor medida se siembra trigo (*Triticum aestivum* L.) como cultivo de invierno.

2.2 Datos a campo

Se relevaron, durante la campaña agrícola 2007/2008, 33 lotes cultivados con soja y maíz. En los mismos, se estableció el cultivo realizado en la campaña anterior a través del análisis del rastrojo muestreado, a los fines de considerar el “tipo de rastrojo” como clase de cobertura para realizar las clasificaciones. A su vez se registró el cultivo de la campaña en estudio. Estas dos variables se utilizaron como verdad de terreno para la aplicación de los métodos supervisados.

2.3 Imágenes satelitales y datos de cobertura

Se utilizaron dos imágenes satelitales Landsat, una tomada el 20/12/2007, correspondiente al inicio de la temporada, y la otra con fecha del 01/03/2008 en la que los cultivos se encontraban en etapas avanzadas de su ciclo. En cada imagen se hizo un recorte con el área de estudio y sobre éste se realizaron las clasificaciones.

Para clasificar la imagen del 20/12/2007 se consideraron tres grupos que abarcaban las clases de cobertura con diferentes criterios.

El primer grupo, G1, incluyó los cultivos de la campaña (soja y maíz) combinados con los tipos de rastrojo encontrados (soja, maíz y trigo). Quedaron establecidas entonces:

Cob 1: Maíz temprano¹ con rastrojo de soja.

Cob 2: Soja temprana con rastrojo de soja.

Cob 3: Soja temprana con rastrojo de maíz.

Cob 4: Soja temprana con rastrojo de trigo.

Cob 5: Sin cultivo con rastrojo de maíz.

Cob 6: Sin cultivo con rastrojo de trigo.

Cob 7: Sin cultivo con rastrojo de soja.

¹ Se consideran como cultivos tempranos a aquellos cuya emergencia fue anterior al 12/12/2007.

En el segundo grupo, G2, se construyeron las clases teniendo en cuenta como única variable el tipo de rastrojo presente, quedando las siguientes coberturas:

Cob 1: Rastrojo de trigo, con soja temprana o sin cultivo.

Cob 2: Rastrojo de maíz, con soja temprana o sin cultivo.

Cob 3: Rastrojo de soja, con maíz temprano, soja temprana o sin cultivo.

En el grupo G3 las clases de cobertura fueron agrupadas considerando como única variable al cultivo de la campaña en estudio, quedando las siguientes:

Cob 1: Maíz temprano con rastrojo de soja.

Cob 2: Soja temprana con rastrojo de soja, maíz y/o trigo.

Cob 3: Sin cultivo con rastrojo de soja, maíz y/o trigo.

En la imagen del 01/03/2008 para la clasificación se consideraron:

Grupo G4, con las coberturas que surgen de combinar los cultivos de la campaña (soja y maíz) con su respectiva época de siembra:

Cob 1: Soja temprana.

Cob 2: Soja tardía².

Cob 3: Maíz temprano.

Cob 4: Maíz tardío.

En el grupo G5 solo se tuvo en cuenta el cultivo implantado, quedando las siguientes clases:

Cob 1: Soja.

Cob 2: Maíz.

2.4 Algoritmos de clasificación supervisada

Las áreas de estudio fueron clasificadas utilizando los siguientes algoritmos: Distancia mínima (Dist_min), Distancia de Mahalanobis (Dist_Mah), Máxima probabilidad (Max_pro) y Mapeador del ángulo espectral (MAE). Quedando conformadas 20 escenas clasificadas, (cinco grupos de coberturas por cuatro algoritmos de clasificación) 12 correspondientes a la imagen 1 y 8 a la imagen 2. Todos estos algoritmos fueron ejecutados utilizando el software ENVI 4.6.

2.5 Análisis estadístico:

La exactitud de las clasificaciones fue evaluada mediante un análisis de las matrices de confusión, siendo éste el método más comúnmente utilizado para validar este tipo de modelos. Dicha matriz contiene información sobre los valores totales o porcentajes de datos observados y estimados para cada característica o propiedad que se

² Cultivos tardíos: se considera al que fue sembrado con posterioridad al 12/12/2007.

clasifique, y a partir de ésta se pueden calcular distintos parámetros que indican la precisión de la clasificación [13].

Los parámetros que surgen de la matriz de errores y la formula para calcularlos son:

Precisión global (PG):

$$P = \frac{\sum_{i=1}^m x_{ii}}{N} \quad (1)$$

Precisión del productor (Pp):

$$P_p = \frac{x_{ii}}{x_{\Sigma i}} \quad (2)$$

Precisión del usuario (Pu):

$$P_p = \frac{x_{ii}}{x_{i\Sigma}} \quad (3)$$

Coefficiente Kappa (κ):

$$\kappa = \frac{N \sum_{i=1}^m x_{ii} - \sum_{i=1}^m x_{i\Sigma} x_{\Sigma i}}{N^2 - \sum_{i=1}^m x_{i\Sigma} x_{\Sigma i}} \quad (4)$$

Donde m = Número total de clases, N = Número total de píxeles en las m clases de referencia, x_{ii} = elementos de la diagonal principal de la matriz de confusión, $x_{\Sigma i}$ = suma de los píxeles de la clase i de referencia y $x_{i\Sigma}$ = suma de los píxeles clasificados como la clase i .

Para cada escena clasificada se generó una matriz de confusión, utilizada para examinar los errores de comisión y de omisión, la exactitud del productor, del usuario y de la clasificación en general. A su vez se calculó el índice kappa (κ), el cual computa el acuerdo entre la imagen clasificada y la realidad observada debido únicamente a la exactitud de la clasificación, suprimiendo el acuerdo que cabría esperar simplemente por azar. Para la interpretación del valor del índice κ , Monserud y Leemans [14] propusieron el uso de una escala donde valores menores a 40% se clasifican como pobres, 40-55% suficientes, 55-70% buenos, 70-85% muy buenos y mayores a 85% excelentes.

3 Resultados y discusión

En la Tabla 1 se muestran los valores de los índices κ y las precisiones globales (en porcentaje) para cada grupo de cobertura y algoritmo de clasificación.

Tabla 1. Estadísticos de Clasificación de Imágenes Landsat para los distintos algoritmos y agrupaciones de cobertura de suelo.

Imagen	Grupo	Algoritmo de Clasificación	Índice Kappa (κ)	Precisión Global (%)
1	G1	Dist_min	0,48	55,9
		Dist_Mah	0,56	62,9
		Max_pro	0,89	91,1
		MAE	0,53	59,7
	G2	Dist_min	0,60	73,2
		Dist_Mah	0,72	81,9
		Max_pro	0,83	88,8
		MAE	0,44	59,0
	G3	Dist_min	0,50	68,0
		Dist_Mah	0,64	76,6
		Max_pro	0,73	82,9
		MAE	0,66	78,1
2	G4	Dist_min	0,51	67,0
		Dist_Mah	0,51	67,4
		Max_pro	0,69	79,7
		MAE	0,48	63,8
	G5	Dist_min	0,85	95,2
		Dist_Mah	0,93	97,8
		Max_pro	0,94	98,0
		MAE	0,60	82,1

A partir de los resultados, se observa que el algoritmo de máxima probabilidad presenta los mayores valores de los estadísticos que validan los mismos, en todos los grupos, tanto para κ como para la PG.

Analizando las diferencias de valores de Max_Pro entre las coberturas de la imagen 1, se presentan los mayores valores en el G1, donde la mayor cantidad de clases, que consideran el cultivo presente y el rastrojo anterior, permiten una mejor clasificación.

Comparando los valores de los grupos G2 y G3 clasificados con Max_Pro, se puede afirmar que en los estados iniciales de los cultivos el algoritmo clasificador se desempeña mejor considerando el rastrojo (G2) que considerando el cultivo (G3). Esta característica se repite para los clasificadores Dist_Min y Dist_Mah y no para MAE, el cual presenta mejor desempeño cuando se agrupan las clases de cobertura de acuerdo al cultivo presente (G3) y no al rastrojo (G2).

En la Figura 1 se muestra el resultado de la clasificación por Max_pro de la imagen 1, realizada utilizando las clases de cobertura definidas en el G1. Esta clasificación fue la que mejores estadísticos presentó, para dicha imagen.

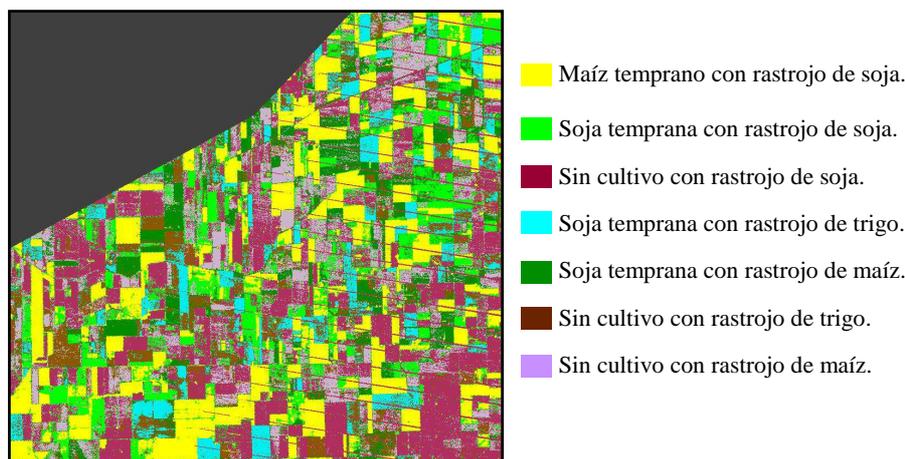


Fig. 1. Recorte de Imagen Landsat (20/12/2007) clasificado por el algoritmo de Máxima de probabilidad, zona central de Córdoba Argentina.

Para la imagen 2, los valores de κ y precisiones globales, para los cuatro clasificadores, son mayores para G5 que para G4, lo que indica que los mismos, cuando los cultivos ya se encuentran en etapas reproductivas (grano maduro para los lotes de maíz temprano, floración para maíz tardío, grano pastoso para soja temprana e inicio de llenado de grano para soja tardía), no son sensibles a la fecha de siembra, y presentan una alta precisión al clasificar solamente el tipo de cultivo (soja y maíz).

Corts *et al.* [1] en un trabajo para clasificar, a partir de MODIS, cobertura terrestre a escala regional, también comprobaron que disminuyendo el número de clases, es decir juntando las clases con mayor dificultad de discriminación, se obtenía mejor precisión en la clasificación. Estos autores afirman que en función de los requerimientos el usuario debe decidir si da prioridad al número de clases o a la precisión final de la clasificación, aseveración que compartimos.

El menor valor de κ encontrado para el algoritmo de clasificación Max_Pro fue para G4, no obstante el mismo es considerado como bueno en la escala propuesta. Este estadístico, para G2 y G3 presenta un valor muy bueno, y para G1 y G5 la valoración es excelente. Bocco *et al.*, clasificando imágenes MODIS con modelos de redes neuronales, obtuvieron valores de κ en un rango de 0,48 y 0,86 con PG entre 76,2% y 92,9%, cuando se clasificó solamente en tres categorías (soja-maíz-suelo desnudo) [7].

En las clases del G1 un examen individual de la precisión de la clasificación del tipo de cobertura del suelo, muestra que el método de clasificación Max_Pro fue la única rutina que alcanzo valores mayores a 85% de precisión de clasificación, tanto productor como usuario (excepto para sin cultivo con rastrojo de maíz, en precisión del usuario). En las clases del G5 se alcanzaron valores categorizados como excelentes para precisión de productor y usuario en todos los métodos, excepto el

MAE. Con este último clasificador los resultados obtenidos se categorizan, no obstante, como muy buenos. En la Tabla 2 se muestran los valores de P_p y P_u de los cuatro algoritmos de clasificación, para las clases de cada grupo.

Tabla 2. Valores (en %) de los estadísticos de Precisión del Productor (P_p) y del Usuario (P_u) para los distintos algoritmos y agrupaciones de cobertura de suelo.

Grupo	Tipo de Cobertura	Dist_min		Dist_Mah		Max_pro		MAE	
		P_p	P_u	P_p	P_u	P_p	P_u	P_p	P_u
G1	Cob 1	99,8	99,8	98,5	100	100	100	99,0	96,0
	Cob 2	52,5	85,5	60,0	49,4	81,0	94,4	51,1	95,8
	Cob 3	66,2	48,4	71,7	51,7	96,3	88,1	78,3	50,2
	Cob 4	94,6	33,5	85,7	39,5	98,9	94,6	90,3	29,3
	Cob 5	4,4	6,0	27,9	41,1	90,3	72,2	17,7	31,3
	Cob 6	19,4	84,1	36,0	90,6	88,6	98,9	11,6	80,0
	Cob 7	71,5	42,8	66,7	51,6	80,5	93,2	87,3	51,4
G2	Cob 1	74,1	89,3	74,2	99,8	90,5	98,3	74,2	95,7
	Cob 2	92,4	51,8	95,1	62,0	97,7	72,6	97,1	51,3
	Cob 3	61,1	99,9	78,0	96,4	82,5	99,2	28,0	100
G3	Cob 1	100	99,0	99,7	99,8	99,8	100	99,3	93,5
	Cob 2	52,2	44,3	70,9	56,8	73,7	68,0	54,0	78,2
	Cob 3	61,9	69,4	69,0	80,4	80,1	84,1	82,0	78,2
G4	Cob 1	49,0	75,7	62,7	62,3	83,8	69,3	43,2	80,7
	Cob 2	67,4	62,7	56,1	62,8	66,3	81,9	65,3	59,6
	Cob 3	100	89,0	100	99,2	100	100	99,4	100
	Cob 4	100	18,0	100	29,8	100	100	100	95,6
G5	Cob 1	96,2	97,8	99,4	97,8	97,6	99,9	81,9	100
	Cob 2	91,2	85,7	91,3	97,6	99,7	91,2	82,6	100

Guerschman *et al.* utilizando datos multi-temporales de Landsat TM y algoritmo clasificador de Máxima probabilidad, al clasificar cobertura del suelo en 14 clases que incluían maíz y soja, obtuvieron valores de $\kappa = 0,55$ y $PG = 62,6\%$; con $P_p = 65\%$ y $P_u = 23\%$ para maíz y $P_p = 74\%$ y $P_u = 83\%$ para soja; y al disminuir la cantidad de clases de la clasificación, sus resultados mejoraron a valores de $\kappa = 0,84$ y $PG = 88,7\%$; con P_p y $P_u = 79\%$ para cultivos de verano [6]. Por otro lado Vizzotto Cattani *et al.* clasificaron en tres clases de cobertura (soja, maíz y otros) imágenes de los satélites RapidEye, de una zona del Estado de Paraná (Brasil), utilizando los mismos algoritmos que en el presente trabajo, y obtuvieron valores de κ entre 0,61 y 0,86 y de PG entre 74 y 91%; concluyendo que el clasificador de Máxima probabilidad se desempeñó mejor que los otros, lo que es coincidente con lo que encontramos en el presente trabajo [10].

4 Conclusiones

Es posible diferenciar tipos de cobertura del suelo agrícola, para el área central de Córdoba, por medio de clasificaciones supervisadas a partir de imágenes del satélite Landsat TM. La clasificación con distintos niveles de desagregación de cultivos y/o residuos alcanza índices muy buenos de precisión.

De los cuatro algoritmos evaluados, el de Máxima Probabilidad proporciona los mejores resultados, tanto en las etapas iniciales del cultivo, como en avanzadas. En la imagen que corresponde a las primeras etapas del ciclo del cultivo, fue posible diferenciar los cultivos presentes y también el tipo de rastrojo remanente de la campaña anterior.

Agradecimientos. Los autores expresan su gratitud a la Secretaría de Ciencia y Técnica – Universidad Nacional de Córdoba - Argentina (Secyt-UNC) por el apoyo financiero que permitió la realización de este trabajo.

Referencias

1. Corts, R., Navarro, M., Pardo, J.A., Bisbal, M.: Clasificación Periódica de Coberturas Terrestres a Escala Regional con Imágenes Modis. *Revista Internacional de Ciencia y Tecnología de la Información geográfica GeoFocus*. 10, 1–17 (2010).
2. Wardlow, B.D., Egbert, S.L., Kastens, J.H.: Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sensing of Environment*. 108, 290–310 (2007).
3. Foody, G.M.: Status of land cover classification accuracy assessment. *Remote Sensing of Environment*. 80, 185–201 (2002).
4. Pal, M., Mather, P.M.: Assessment of the effectiveness of support vector machines for hyperspectral data. *Future Generation Computer Systems*. 20, 1215–1225 (2004).
5. Rizzi, R., Rudorff, B.F.T.: Estimativa da área de soja no Rio Grande do Sul por meio de imagens Landsat. *Revista Brasileira de Cartografia* 57, 226–234 (2005).
6. Guerschman, J.P., Paruelo, J.M., Di Bella, C., Giallorenzi, M.C., Pacin, F.: Land cover classification in the Argentine Pampas using multi-temporal Landsat TM data. *Int. J. Remote Sens.* 24, 3381–3402 (2003).
7. Bocco, M., Ovando, G., Sayago, S., Willington, E.: Neural Network Models for Land Cover Classification from Satellite Images. *Agricultura Técnica*. 67, 414–421 (2007).
8. South, S., Qi, J., Lusch, D.P.: Optimal classification methods for mapping agricultural tillage practices. *Remote Sensing of Environment*. 91, 90–97 (2004).
9. Castillejo-González, I.L., López-Granados, F., García-Ferrer, A., Peña-Barragán, J.M., Jurado-Expósito, M., Sánchez de la Orden, M., González-Audicana, M.: Object- and pixel-Based Analysis for Mapping Crops and their Agro-Environmental Associated Measures Using QuickBird Imagery. *Computers and Electronics in Agriculture*. 68, 207–215 (2009).
10. Vizzotto Cattani, C.E., Mercante, E., Wachholz de Souza, C.H., Costa Wrublack, S.: Desempenho de Algoritmos de Classificação Supervisionada para Imagens dos Satélites RapidEye. En XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR, Foz do Iguaçu, PR. pp 8005–8010. Brasil, (2013).

11. Ministerio de Agricultura, Ganadería y Pesca de la Nación, <http://www.siiia.gov.ar/index.php/series-por-tema/agricultura>. Consultado en abril 2013
12. Bocco, M., Ovando, G., Sayago, S., Wellington, E., Heredia, S.: Estimating Soybean Ground Cover from Satellite Images Using Neural Networks Models. *International Journal of Remote Sensing*. 33, 1717–1728 (2012).
13. Ruiz Posse, E., Bocco, M. Bisonard, M.: Técnicas de Geoprocesamiento en el Estudio de Enfermedades Transmitidas por Vectores en Maíz. En: *Enfermedades del Maíz Producidas por Virus y Mollicutes en Argentina*. Ed. Gimenez Pecci, M., Laguna, I., Lenardon, S. INTA 2006. pp. 179–190.
14. Monserud, R.A., Leemans, R.: Comparing Global Vegetation Maps with the Kappa Statistic. *Ecological Modelling*. 62, 275–293 (1992).